

Final Report:

An Acoustic Charge Transport Imager for High Definition Television

by

Dr. William D. Hunt

Dr. Kevin Brennan

Dr. Gary May

School of Electrical and Computer Engineering
Georgia Institute of Technology

with assistance from:

Dr. William E. Glenn

Florida Atlantic University

Mr. Mike Richardson

Eastman Kodak Company

Dr. Richard Solomon

Massachusetts Institute of Technology/Polaroid Corporation

For GIT Project E-21-687

November 5, 1999

EXECUTIVE SUMMARY

Executive Summary

This project, over its term, included funding to a variety of companies and organizations. In addition to Georgia Tech these included Florida Atlantic University with Dr. William E. Glenn as the P.I., Kodak with Mr. Mike Richardson as the P.I. and M.I.T./Polaroid with Dr. Richard Solomon as the P.I. The focus of the work conducted by these organizations was the development of camera hardware for HDTV. The focus of the research at Georgia Tech was the development of new semiconductor technology to achieve a next generation solid state imager chip that would operate at a high frame rate (170 frames per second), operate at low light levels (via the use of avalanche photodiodes as the detector element) and contain 2 million pixels. The actual cost required to create this new semiconductor technology was probably at least 5 or 6 times the investment made under this program and hence we fell short of achieving this rather grand goal. We did, however, produce a number of spin-off technologies as a result of our efforts. These include, among others, improved avalanche photodiode structures, significant advancement of the state of understanding of ZnO/GaAs structures and significant contributions to the analysis of general GaAs semiconductor devices and the design of Surface Acoustic Wave resonator filters for wireless communication. More of these will be described in the report.

The work conducted at the partner sites resulted in the development of 4 prototype HDTV cameras. The HDTV camera developed by Kodak uses the Kodak KAI-2091M high-definition monochrome image sensor. This progressively-scanned charge-coupled device (CCD) can operate at video frame rates and has 9 μm square pixels. The photosensitive area has a 16:9 aspect ratio and is consistent with the "Common Image Format" (CIF). It features an active

image area of 1928 horizontal by 1084 vertical pixels and has a 55% fill factor. the camera is designed to operate in continuous mode with an output data rate of 5MHz, which gives a maximum frame rate of 4 frames per second. This camera was delivered to:

Dr. John C. Wang
Deputy Chief, Applied Information Technology Division
NASA Ames Research Center
Bldg. 211, Room 136
Moffett Field, CA 94035
Phone: 650-604-5525
email: jwang@mail.arc.nasa.gov

The MIT/Polaroid group developed two cameras under this program. The cameras have effectively four times the current video spatial resolution and at 60 frames per second are double the normal video frame rate. But spatial and temporal resolution are not enough to create an illusion represented reality: virtually noise-free, coherent, and progressive image capture make this camera design different from all existing commercial video devices. The camera system implements the 1280 x 720 image format supporting the 750/60/1:1 production standard. The 1" video format, 16:9 aspect ratio progressive scan, frame-transfer CCD sensor with square pixels was designed, and sensor incorporation and camera adaptations were implemented. The first prototype of this camera was demonstrated at the 1996 National Association of Broadcasters (NAB) show. Both of these cameras were delivered to:

Dr. Tice DeYoung
NASA Headquarters, Code PT
Washington, DC 20546
Phone: (202) 358-1363
email: tdeyoung@mail.arc.nasa.gov

One of the cameras is currently assigned to the supercomputer center at the Naval Research Laboratory in Anacostia where it is available for further demonstration. Additional cameras have been purchased by ABC/Disney and other commercial firms. Demonstrations have been made at

the National Institutes of Health, the U.S. Army, the National Security Agency, the National Association of Broadcasters annual meeting in Las Vegas (where it won Best of Show), and at the American Society of Cinematographers' annual convention in Los Angeles.

Florida Atlantic University under this program developed a breadboard high definition camera mounted on an optical bench. They used two experimental CCD sensors made by Eastman Kodak in the camera. This was the first demonstration in the world of a color camera with 1920 x 1080 pixels progressively scanned at 60 frames per second. This is the progressive SMPTE 274M standard. It has been a long time objective for HDTV for 15 years. This is the first camera to achieve that objective. Since this was a breadboard using an early version of the Kodak CCD, it was not a practical camera since it could not be moved. The Florida Atlantic University group subsequently built a more practical camera, on a DARPA contract, using the equipment purchased on this subcontract, plus additional materials. This camera was delivered to the Navy for underwater experiments. Work has continued with enhancements on this camera with DARPA funding and with some commercial funding.

SECTION I:

INTRODUCTION AND BACKGROUND:

GEORGIA TECH PORTION OF THE PROGRAM

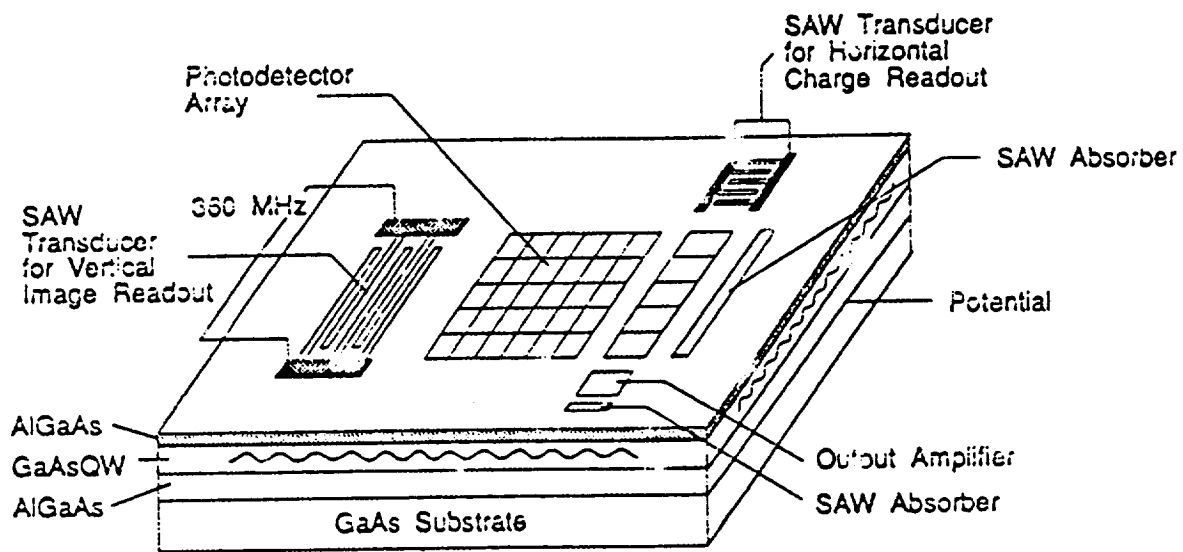
1.0 Introduction and Background: Georgia Tech Portion of the Program

In this project we sought to develop a new kind of solid state imager for HDTV applications. At the inception of this program there were no US manufacturers of HDTV cameras or camera chips and we sought to fill that void. At the time there were also no Japanese or European manufacturers which could supply progressive-scan HDTV camera chips though it was felt strongly that such chips would probably be required by the eventual FCC standard for HDTV. Our rather ambitious goal was to create a new technology for imager chips which would operate at the high frame rates needed. We estimated that with our approach an imager consisting of a 1920 (V) by 1080 (H) array of pixels would have a rate of 170 frames per second when progressively scanned. The pixel size would be roughly 8 μm by 8 μm . Our idea, covered in US Patent No. 5,162,885 "Acoustic Charge Transport Imager", was to combine an avalanche photodetector with acoustic charge transport (ACT) readout. ACT is a charge transport mechanism for which the photogenerated charge can be read out via a voltage which propagates with a surface acoustic wave (SAW) in a piezoelectric semiconductor. Avalanche photodetectors play a role in camera tubes such as the HARPICON H4318 from Hitachi for which the detection stage is an amorphous selenium target under reverse bias. In addition, some CCD imagers utilize amorphous silicon avalanche photodiodes. The use of avalanche photodetectors introduces some optical gain in the chip which is critical for HDTV imaging at moderate light levels.

To be sure, this was an ambitious project. ACT technology was in a relatively early stage of development, having not yet reached commercialization and even GaAs technology was not at a high level of maturity. Though we did not succeed in developing the HDTV ACT imager chip, we made a number of significant discoveries and advances that have had substantial impact on GaAs APDs, GaAs electronics, SAW resonator filters and Zinc Oxide film technology.

The situation has somewhat changed now. GaAs technology, driven by the wireless communications industry has reached maturity and is going well. It is the fastest growing sector of the semiconductor industry with such companies as RF Micro-Devices (RFMD) performing very well. In the last year, since its GaAs Heterostructure Bipolar Transistor (HBT) fabrication facility has come on line, RFMD's stock has risen from roughly \$6 per share to \$86 per share. The path chosen towards an ACT imager chip would be very different if the project were starting today. One would leverage the existing industrial base in both GaAs APDs for fiber optic telecommunications and GaAs integrated circuits. More will be said about this later in the section on suggestions for future work.

In the figure shown below the schematic of the ACT imager architecture is shown. There come to mind several approaches to breaking down this problem. One focal point is the APDs and the work on these is described primarily in Sections 2 and 3 but also to some extent in Section 4. The other focal point is the ACT portion of the device and work on this is presented in Sections 4 and 5.



SECTION II:
AVALANCHE PHOTODIODES

**EXPERIMENTAL & THEORETICAL CHARACTERIZATION OF
GaAs/AlGaAs AVALANCHE PHOTODIODES FOR HDTV SENSOR
APPLICATIONS**

FINAL REPORT

**Dr. Chris Summers
Dr. Hicham Menkara**

**Electro-Optics, Environment & Materials Laboratory
Georgia Tech Research Institute
Atlanta, Ga 30332**

April 15, 1998

TABLE OF CONTENTS

<u>1. GAIN/NOISE PROPERTIES OF DOPED MQW JUNCTIONS</u>	2
<u>1.1 INTRODUCTION</u>	2
<u>1.2 GAIN ENHANCEMENT</u>	2
<u>1.3 DARK CURRENT REDUCTION</u>	13
<u>1.4 EXCESS NOISE REDUCTION</u>	16
<u>1.4.1 Johnson Noise</u>	16
<u>1.4.2 Shot Noise</u>	17
<u>1.4.3 Flicker Noise</u>	17
<u>1.4.4 Total Non-multiplication Noise</u>	17
<u>1.4.5 Excess Noise</u>	18
<u>2. SPECTRAL RESPONSE PROPERTIES</u>	23
<u>2.1 APD QUANTUM EFFICIENCY</u>	23
<u>2.2 SPECTRAL RESPONSE DATA</u>	25
<u>3. TIME RESPONSE CHARACTERISTICS</u>	29
<u>3.1 BANDWIDTH LIMITATIONS OF PHOTODETECTORS</u>	30
<u>3.2 OVERALL PHOTODETECTOR BANDWIDTH</u>	32
<u>3.3 EXPERIMENTAL RESULTS</u>	32
<u>4. TEMPERATURE DEPENDENCE</u>	41
<u>4.1 IMPACT IONIZATION VS. TUNNELING</u>	41
<u>4.2 JUNCTION CAPACITANCE AND TEMPERATURE</u>	44
<u>5. EFFECT OF VARIATIONS IN THE DOPING PROFILES</u>	47
<u>5.1 INTRODUCTION</u>	48
<u>5.2 THEORETICAL RESULTS</u>	49
<u>5.3 EXPERIMENTAL RESULTS</u>	57
CONCLUSIONS	61
PUBLICATIONS :	64
PRESENTATION :	64
PH.D. THESIS:	64

1. Gain/Noise Properties of Doped MQW Junctions

1.1 Introduction

In this section, a detailed characterization has been made of the external properties of both the PIN and the MQW structures. Comparison of the gain properties at low voltages between the MQW and conventional APDs showed a direct experimental confirmation of a structure-induced carrier multiplication due to interband impact ionization. Similar studies of the bias dependence of the excess noise characteristics show that the low-voltage gain is primarily due to electron ionization in the MQW-APDs, and to both electron and hole ionization in the conventional APDs. For the doped MQW APDs, the average gain per stage was calculated by comparing gain data with depletion width and carrier profile measurements, and was found to vary from 1.03 at low bias to 1.09 near avalanche breakdown. These results are in good agreement with theoretical models developed by Brennan¹ for similar derivatives of the doped MQW APD.

1.2 Gain Enhancement

As was previously mentioned, superlattice multiplication APDs are designed to outperform bulk multiplication APDs by artificially enhanced ionization through the introduction of multiple quantum well layers. This behavior is attributed to the large difference in the conduction and valence-band edge discontinuities at the AlGaAs/GaAs interface. When a “hot electron” enters from the AlGaAs barrier layer into a GaAs well, it abruptly gains an energy equal to the conduction band gap discontinuity, ΔE_c . The effect is that the electron “sees” an ionization energy reduced by ΔE_c with respect to the

threshold energy in bulk GaAs ($E_{th}=2.0$ eV).² Since the impact ionization rate α increases exponentially with decreasing E_{th} , a large increase in the effective α compared to that in bulk GaAs is expected. When the electron enters the next barrier layer, the threshold energy in the AlGaAs material is increased by ΔE_c therefore decreasing the value of α in the AlGaAs. However, since $\alpha_{GaAs} \gg \alpha_{AlGaAs}$, the exponential dependence on the threshold energy results in an increase in the overall average α given by:

$$\alpha_{avg} = (\alpha_{GaAs} + \alpha_{AlGaAs}) / (L_{GaAs} + L_{AlGaAs}) \quad [1-1]$$

where L represents the layer thicknesses.

In contrast, the ionization rate for holes, β , is not increased substantially due to the smaller valence-band discontinuity. This results in a net enhancement in the α/β ratio.³

The APD devices were characterized under both light and dark conditions using current-voltage (I-V), capacitance-voltage (C-V), and noise measurements. Gain curves were calculated from the reverse bias I-V measurements performed as a function of photon flux. Carrier concentrations and depletion width profiles were determined from the C-V data using a one-sided junction approximation.

The I-V measurements were taken in the dark and under HeNe laser illumination. In order to ensure pure electron injection, the laser beam was focused through a microscope objective at the center of the 75 μm diameter opening in the upper p-contact ring. Breakdown voltages, V_B , were measured to be about 27 V for the doped MQW APD and 63 V for the conventional APD with corresponding dark currents, measured at 20% of the breakdown voltages, of about 10 pA and 100 pA, respectively.

The dark current IV plots are shown in Figure 1-1 where the low breakdown voltage characteristics of the doped MQW APD is demonstrated. This is a result of the high doping present in the junction which helps increase the electric field magnitude closer to the its critical avalanche value.

The C-V measurements were performed at 1 MHz. The C-V data (shown in Figure 1-2) was then analyzed to calculate the depletion widths and carrier profiles for the two structures. The net carrier concentration for the conventional APD is shown in Figure 1-3 as a function of the calculated depletion width. As the reverse bias is increased, the capacitance decreases to 0.8 pf, while the depletion width increases to about 2.6 μm prior to breakdown around 63V. Note that the carrier concentration increases sharply as the depletion edge is extended into the doped contact region. Figure 1-4 shows the corresponding plot for the doped-well MQW device which had a breakdown voltage of about 27 Volts . This plot clearly shows evidence of the depletion of all 10 stages in the MQW APD. The presence of the peaks in the carrier profile data is due to the unequal p and n doping concentrations in the wells. This results in partial

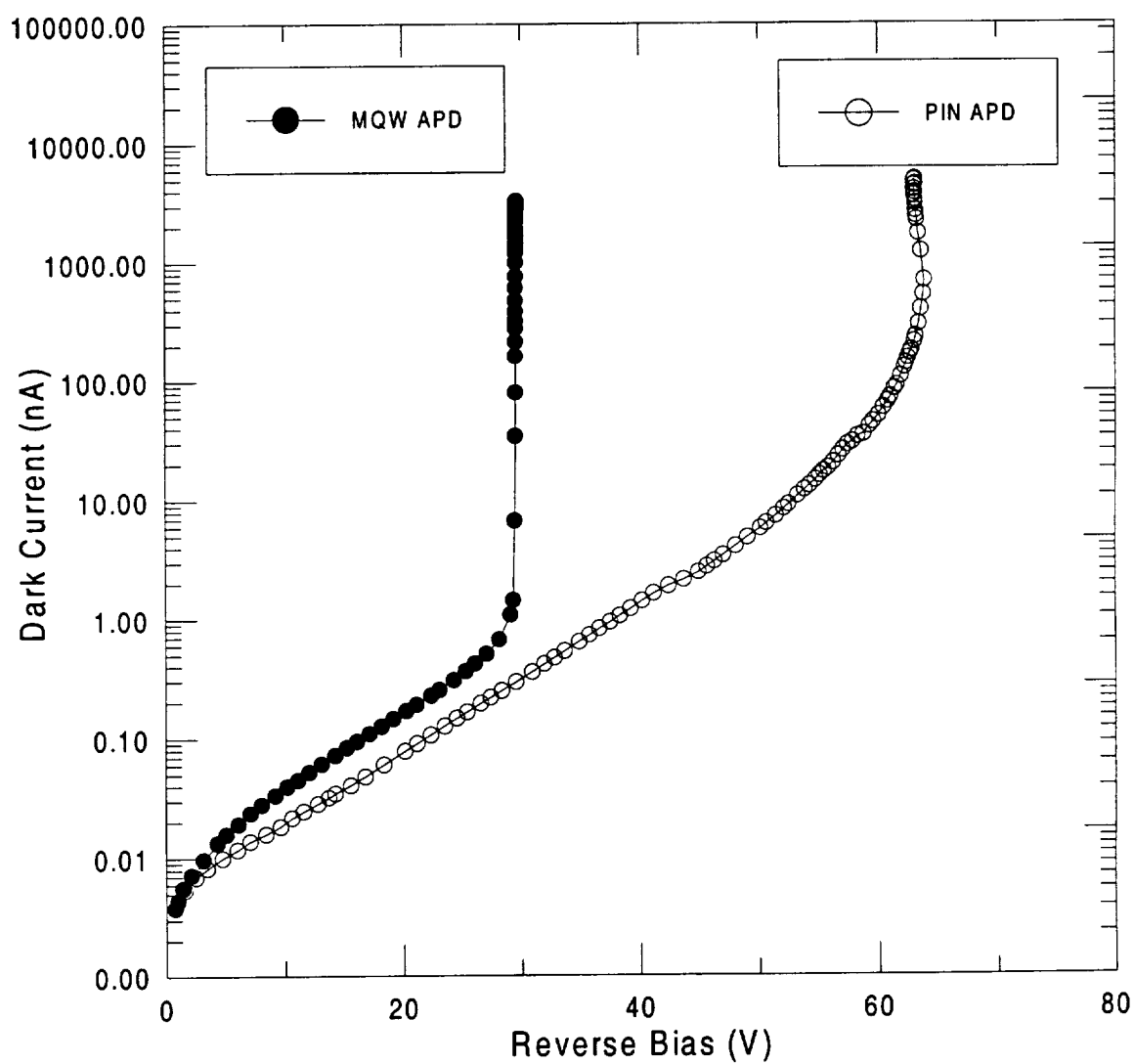


Figure 1-1: Dark IV plots for the PIN and the doped-well MQW APD

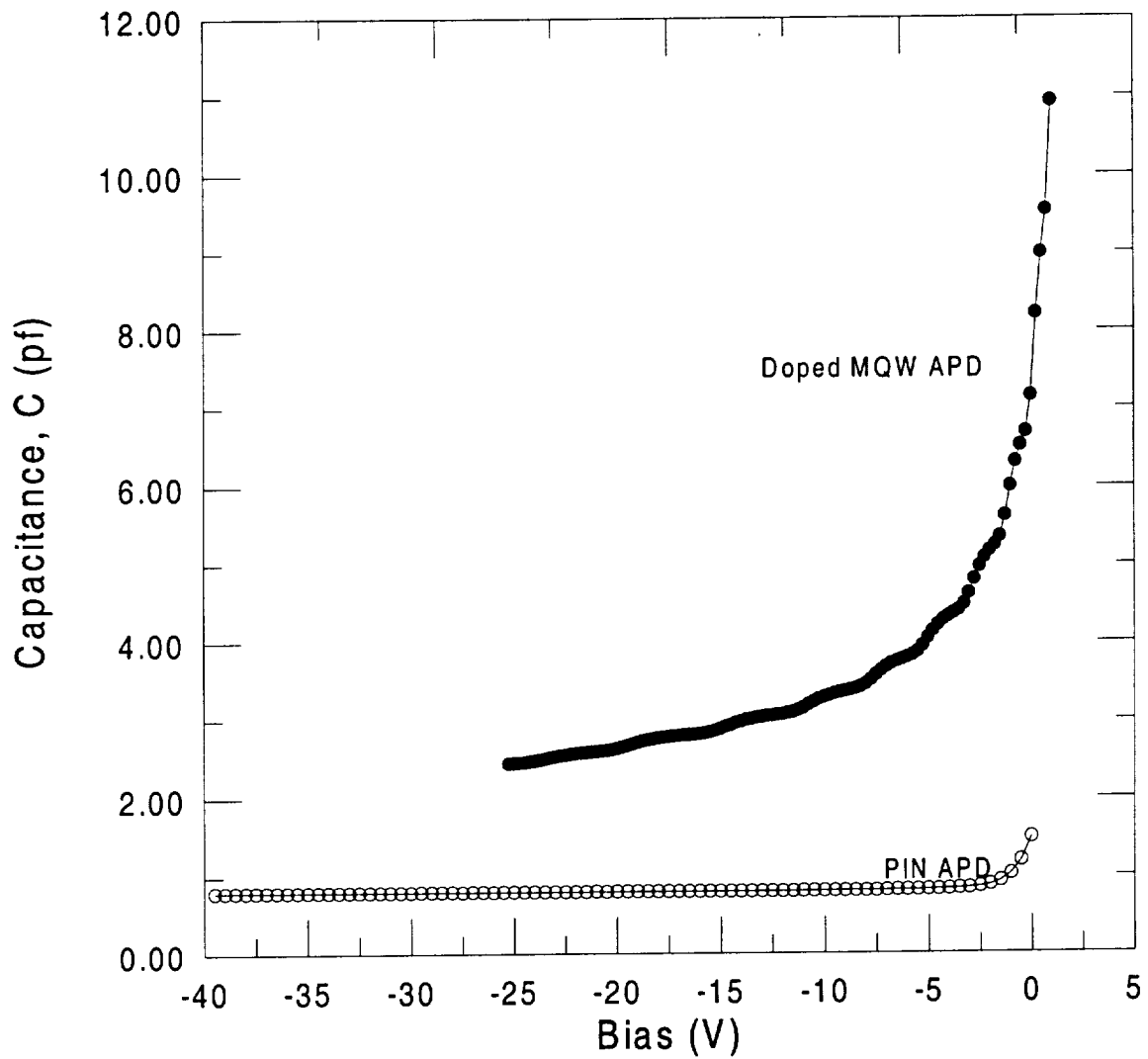


Figure 1-2: CV plots for PIN and a doped MQW APDs

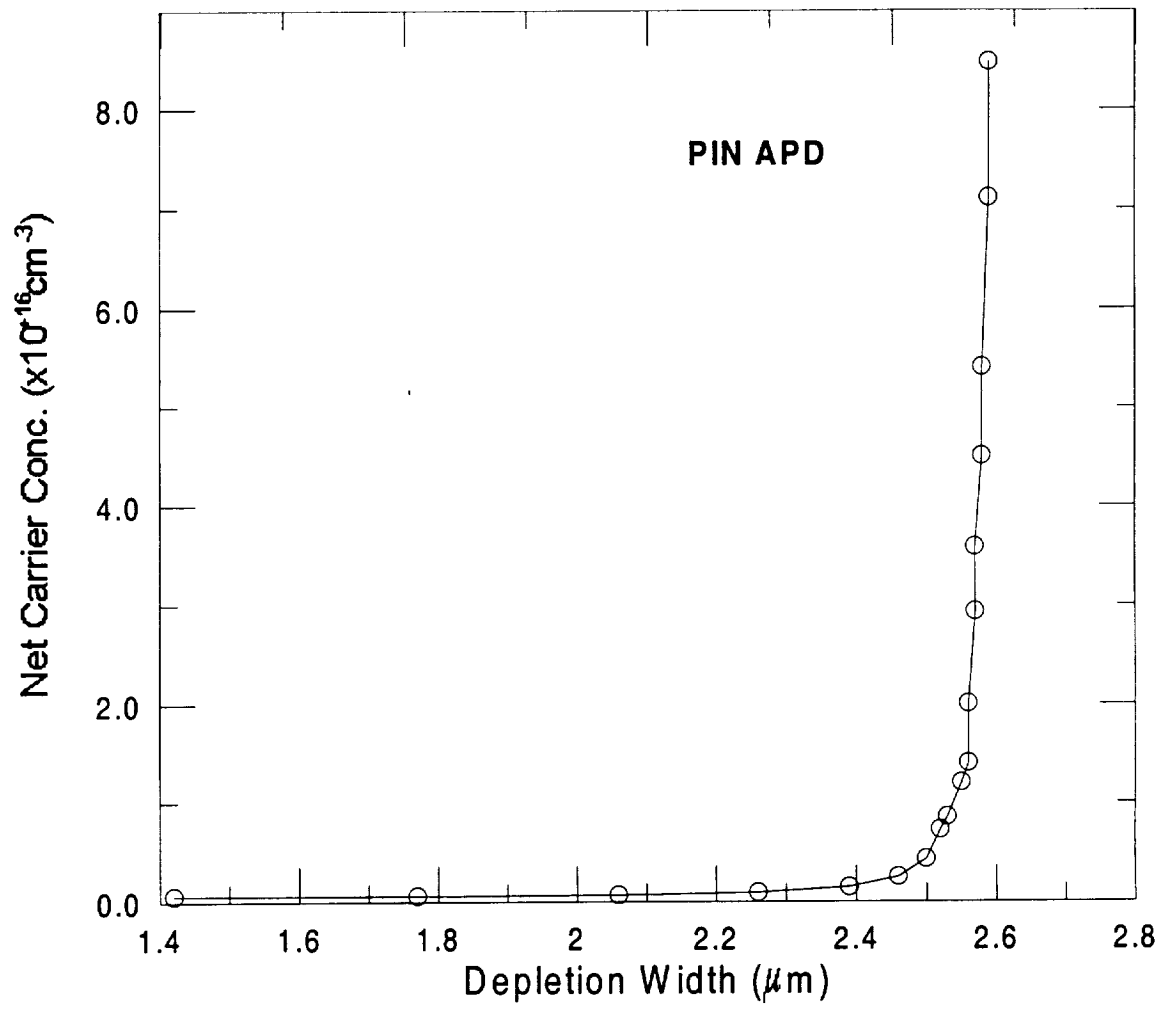


Figure 1-3: Carrier concentration profile vs depletion width for the PIN APD.

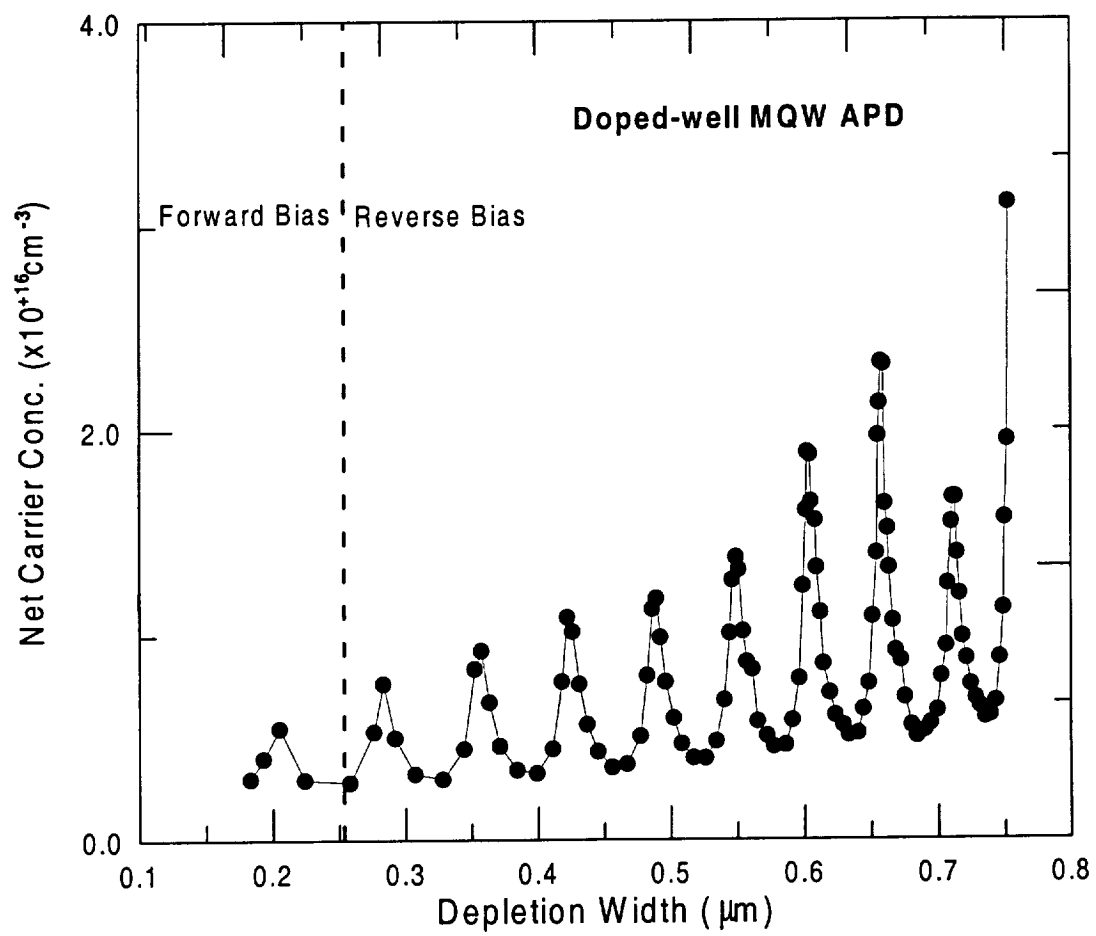


Figure 1-4: Carrier concentration profile vs depletion width for the doped MQW APD.

depletion of the junction as will be shown shortly. With applied bias, the depletion width increases, and additional quantum-well stages become depleted. This gives rise to ripples in the CV profile resulting from peaks in the carrier concentration. Note that the peak positions do not agree with the 1000 Å period of the MQW structure. This discrepancy is due to the fact that carrier concentration profiles were calculated assuming a one sided depletion.⁴ This is generally not the case in such structures unless there is a large doping imbalance in the junction preventing it from depleting both ways. In addition, the spatial resolution of the C-V measurements was limited by the Debye length given by⁵,

$$L_D = \sqrt{(kT\epsilon_s / q^2 N)} \quad [1-2]$$

which is about 40 Å at room temperature for a doping level of $n=1.5 \times 10^{18} \text{ cm}^{-3}$. The Debye length is the distance over which the Coulomb (electrostatic) forces between charged layers are essentially screened out. Since the thickness of the doped layers in the wells was of the same order of magnitude (50 Å), abrupt changes in the doping concentration could not be accurately measured.

The gain curves, calculated from the I-V data, are shown in Figure 1-5 where the bias values were normalized by the breakdown voltage of each device to enable comparison. Figure 1-5(a) clearly shows the presence of gain in the doped MQW device in the low voltage region while the conventional p-i-n structure (Figure 1-5(b)) does not show any gain in this regime. This is an indication of a structure-induced carrier multiplication resulting from the band discontinuity and the doping in the MQW APD. In order to calculate the gain per period in the doped MQW, the carrier profile plot (Figure 1-4) was superimposed on the gain curve (Figure 1-5(a)) and the gain was estimated at

each consecutive carrier concentration minimum as shown in Figure 1-6. The corresponding gain values per stage were found to increase from 1.03 at low bias (one depleted stage), to about 1.09 near breakdown (ten depleted stages). These results are in good agreement with theoretical predictions provided by Brennan¹ for similar derivatives of the doped MQW APD.

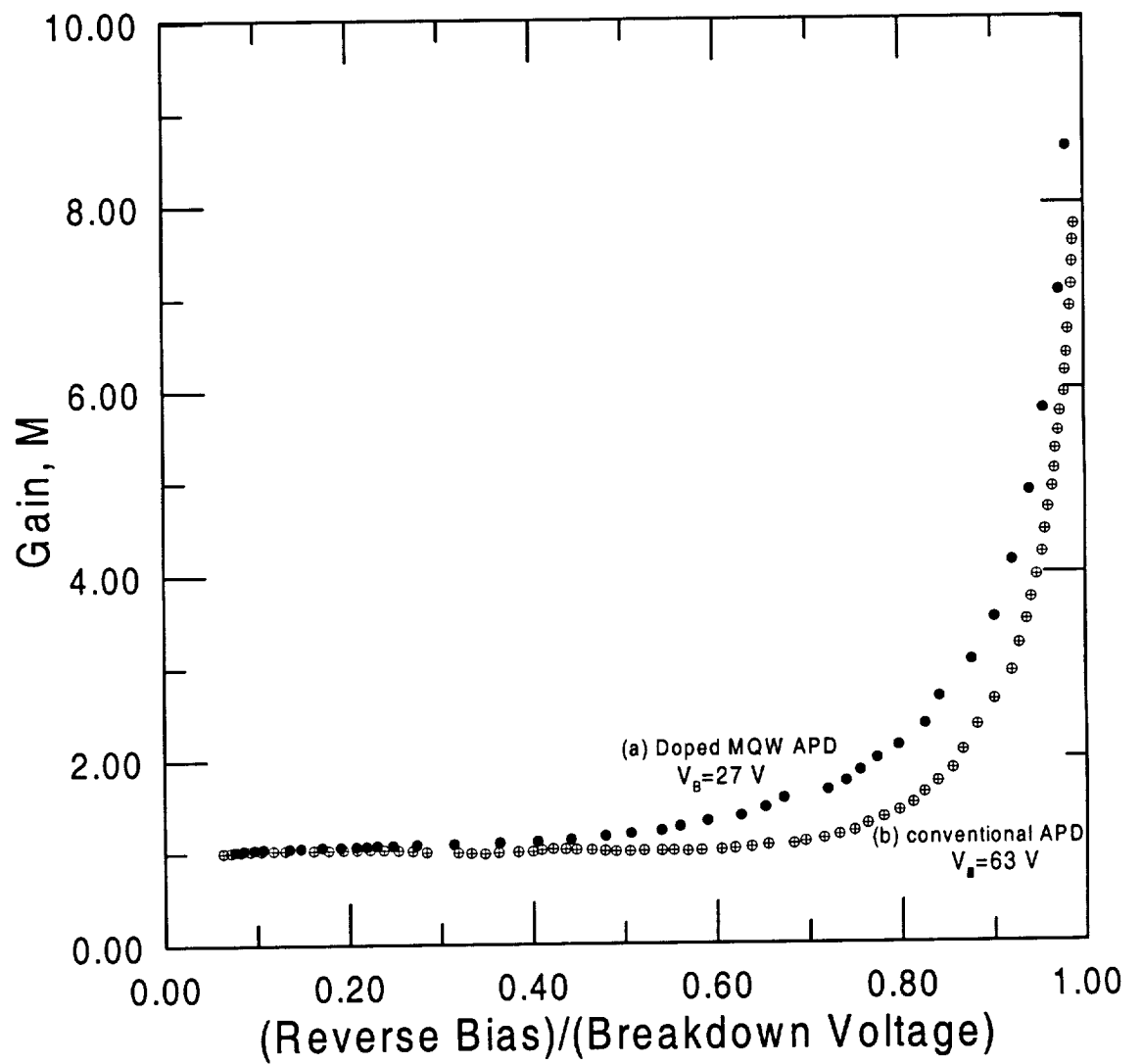


Figure 1-5: Gain vs. the ratio of reverse bias to breakdown voltage for the (a)MQW APD and (b)PIN APD.

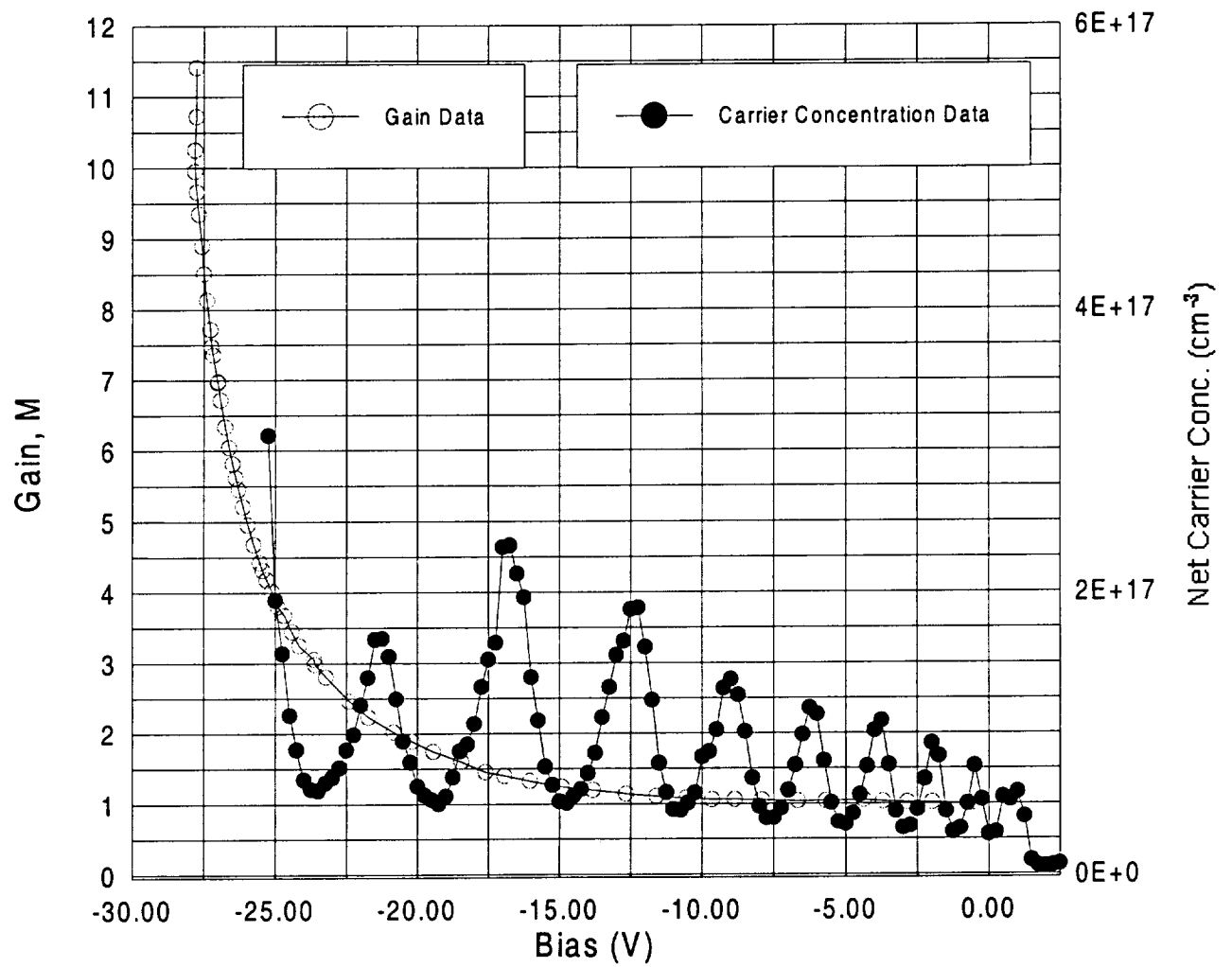


Figure 1-6: Plot used for determining gain per stage for a doped MQW APD

1.3 Dark Current Reduction

Dark current is one of the main parameters of concern in photodetectors. Dark current analysis and the reduction of dark current are very important for high sensitivity and low noise applications. Defect centers, heterojunction interface traps, as well as mesa surface leakage can generate high levels of excess dark current and reduce minority-carrier lifetime.⁶

In a typical PN junction, the overall dark current is the sum of the bulk and surface components. The bulk component is usually made up of diffusion, generation-recombination, and tunneling currents. The surface component consists of generation-recombination, and leakage shunt currents usually formed at semiconductor and dielectric interfaces.⁷ In a device structure with top p and n contacts such as the APDs used in our experiments, there are additional sources of dark current components. These are due to defect centers at the GaAs/AlGaAs interface and most importantly to surface leakage currents along the mesa edge which can contribute significantly to the dark current.

In this section, it will be shown how substantial the surface leakage component can be and how certain growth, processing and surface treatment techniques can be used to dramatically lower surface leakage currents by several orders of magnitude. The devices that were measured were volume- and delta-doped MQW. In the volume-doped MQW structures, the GaAs wells were doped with 50 Å ($3.0 \times 10^{18} \text{ cm}^{-3}$) adjacent p⁺ and n⁺ layers. In the delta-doped APDs, p⁺ and n⁺ layers with a sheet charge density of $1\text{-}5 \times 10^{12} \text{ cm}^{-2}$ were introduced separated by undoped spacer layers ranging from 50 to 150 Å. Through careful dopant calibration, the devices were grown such as to achieve full depletion at low bias. After processing the devices into mesa diodes, various surface

passivation treatments were investigated. These include both plasma ashing in an O_2 plasma and ammonium sulfide treatments. Through the application of such treatments, a decrease in the reverse bias dark current by as much as a factor of 1000 was achieved in the low bias region. This can be seen in Figure 1-7 where the dark current is plotted both before and after surface treatment by ammonium sulfide. The dark current approximately follows a square-root behavior at low to medium reverse bias while at high biases, avalanche currents dominate. The rapid increase in the dark current at low reverse bias and the large drop in its value achieved by surface treatment are indicative of surface leakage. The leakage in these heterojunction mesa diodes was dominated by generation/recombination current near the intersection of the mesa surface with the GaAs/AlGaAs depletion region.

As a result of surface treatment, dark currents as low as 1 pA were obtained under zero applied bias. In some APDs, the dark currents increased to only 12 pA at 20% of breakdown. In addition, these devices exhibited extremely high gains which exceeded 10,000 in some cases. In most traditional APDs, the presence of high dark currents usually presents a limiting factor preventing the further increase in a device's photocurrent gain beyond avalanche breakdown. By reducing the dark currents in these devices, it was possible to maintain it at levels well below that of the photocurrent. This made it possible to achieve and sustain high levels of gains well beyond breakdown.

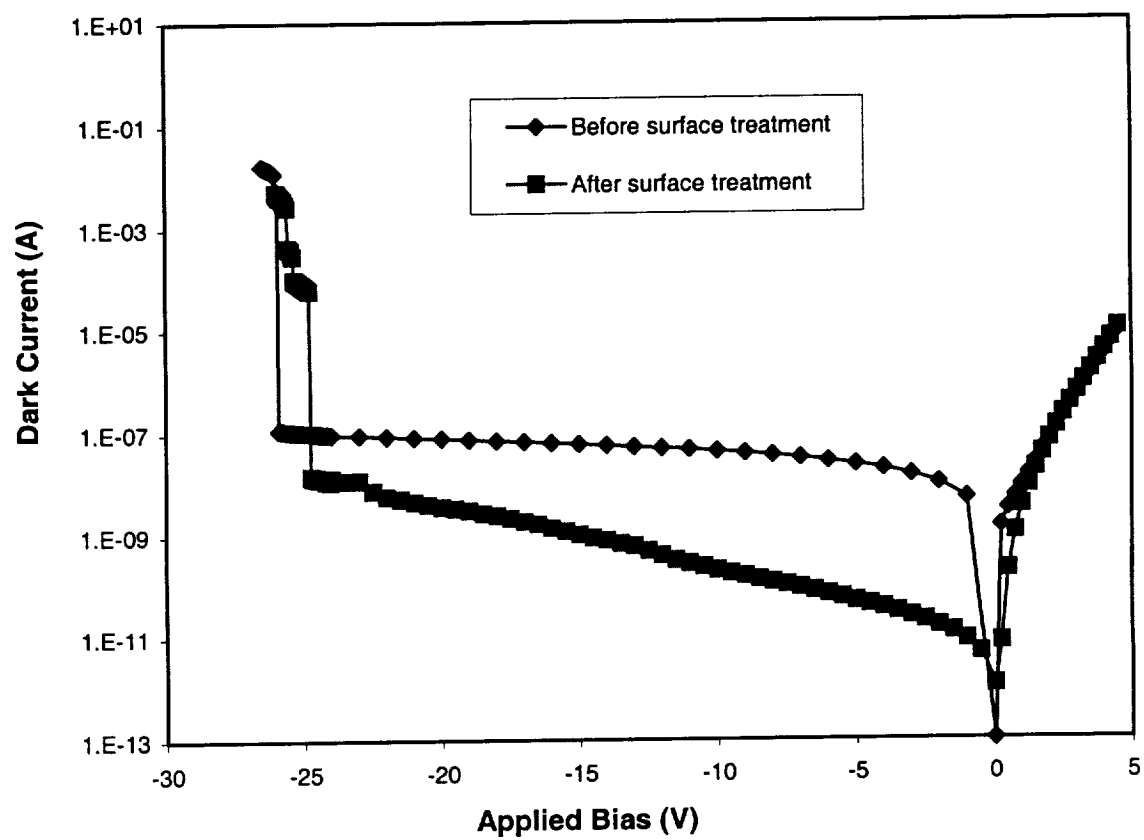


Figure 1-7: Dark current reduction as a result of surface treatment by ammonium sulfide

1.4 Excess Noise Reduction

There are many different types of noise that may be present in an electrical device. A brief discussion of the various types of noise mechanisms is presented below.

1.4.1 Johnson Noise

Johnson noise is caused by the random motion of thermally energetic electrons in resistive materials. Its instantaneous amplitude is not predictable, but the probability of its amplitude being within an interval of dV volts is equal to $p(V)dV$ where $p(V)$ is expressed by the familiar Gaussian probability function:

$$p(V) = \frac{1}{(2\sigma^2)^{1/2}} e^{-V^2/2\sigma^2} \quad [1-3]$$

where the parameter σ is the rms value of the fluctuations and the quantity universally accepted to describe the noise output from a resistor. σ is bandwidth dependent and is expressed as follows:

$$\sigma = (4kTR_s B)^{1/2} \quad (\text{volts}) \quad [1-4]$$

where k is Boltzmann's constant, T is the resistor temperature in K, R_s is the resistance in ohms, and B is the noise bandwidth in hertz. Johnson noise is "white noise", that is the rms value per unit bandwidth (rms density) is constant from DC to frequencies extending into the infrared region.

1.4.2 Shot Noise

Shot noise is the result of random current fluctuations in vacuum tubes and semiconductor junctions. It is caused by the random arrival of discrete electron charges at anodes, collectors, and drains. The rms value of shot noise is given by:

$$I_{shot} = (2eI_{dc}B)^{1/2} \quad (\text{A}) \quad [1-5]$$

where e is the electron charge, I_{dc} is the average DC current through the diode, and B is the noise bandwidth in hertz.

1.4.3 Flicker Noise

Flicker noise is characterized by its spectral composition and for most electronic devices, it dominates thermal and shot noise from DC to about 100 Hz. Although flicker noise can be detected in virtually all conducting materials with applied power, it seems to be most prominent where electron conduction occurs in granular or semiconductor devices. For most semiconductor devices, flicker noise is due to surface effects resulting in random carrier recombinations at interface traps. Flicker noise exhibits a $1/f^n$ power spectrum, with n typically ranging from 0.9 to 1.35.

1.4.4 Total Non-multiplication Noise

Because all the noise sources are considered to be random and uncorrelated, the noise power in a system is additive, and the total rms noise is the square root of the sum of the squares of each of the three noise sources previously described. The total non-multiplication noise output voltage is given by :

$$E_{tno} = [4kTR_sB + (I_{shot}R_s)^2 + e_f^2]^{1/2} \quad \text{volts rms} \quad [1-6]$$

1.4.5 Excess Noise

The excess noise factor is the component of total noise corresponding to fluctuations in the process of carrier multiplication in an avalanche photodiode. It is defined as the ratio of multiplication-related noise to that of the non-multiplication noise defined by equation [5-6]. Excess noise measurements are usually conducted at high frequencies where the shot noise is the most dominant non-multiplication term. Therefore, all other noise processes are commonly ignored in excess noise computations. McIntyre has shown⁸ that the statistical nature of the multiplication process adds an additional component to the noise which can be included with the shot noise of the APD as an excess noise factor. The excess noise factor in the case of pure electron injection is given by:

$$F_e = \frac{M}{k} + (1 - \frac{1}{k})(2 - \frac{1}{M}) \quad [1-7]$$

where M is the multiplication factor, and k is the effective electron to hole ionization ratio of the APD. The root mean square noise current $\langle i_n^2 \rangle$ can be expressed as:

$$\langle i_n^2 \rangle = 2eI_{p0}M^2FB \quad [1-8]$$

where I_{p0} is the primary multiplied photocurrent. In other words, the actual photocurrent is given as:

$$I_{ph} = I_{p0} * M \quad \text{for } I_{ph} \gg I_D \text{ (dark current)} \quad [1-9]$$

A plot of F(M) vs. M from McIntyre's theory is shown in Figure 1-8 for k' ($=1/k=\beta/\alpha$) ranging from 0.001 to 1000. The plots are approximately symmetric on a log-log scale about the axis $F(M)=M$ for k and 1/k. At any given gain, lower excess noise is obtained if

the carrier with the higher ionization coefficient is injected into the multiplication region. For low k' , if the correct carrier is injected, the excess noise can be quite low, with a limiting value of 2 for $k'=0$ at high gain. However, if the wrong carrier is injected, the excess noise becomes very high, with the penalty becoming increasingly more severe as the disparity between ionization coefficients decreases. Thus, it is important to inject the carrier with the higher ionization coefficient into the multiplication region. The lower the k' (or higher the k), the higher the relative difference between α and β , and the lower the excess noise. It is important to note that the McIntyre model is not well suited for describing the noise characteristics of MQW devices since it was intended mainly for conventional APDs. Better models have been developed by Teich et al.^{9,10} and are described in the literature. In addition, Marsland^{11,12} and Hayat¹³ have recently considered the “dead space” between ionization events in their excess noise calculations. They concluded that McIntyre’s calculations overestimate the excess noise factor for a given k . McIntyre curves were used in our excess noise factor plot for comparison purposes in order to clearly illustrate the difference between the noise properties of conventional and MQW APDs. Multiplication noise measurements were conducted on both APDs using an HP8568B spectrum analyzer set at a 200 kHz center frequency with a 10 kHz resolution bandwidth.

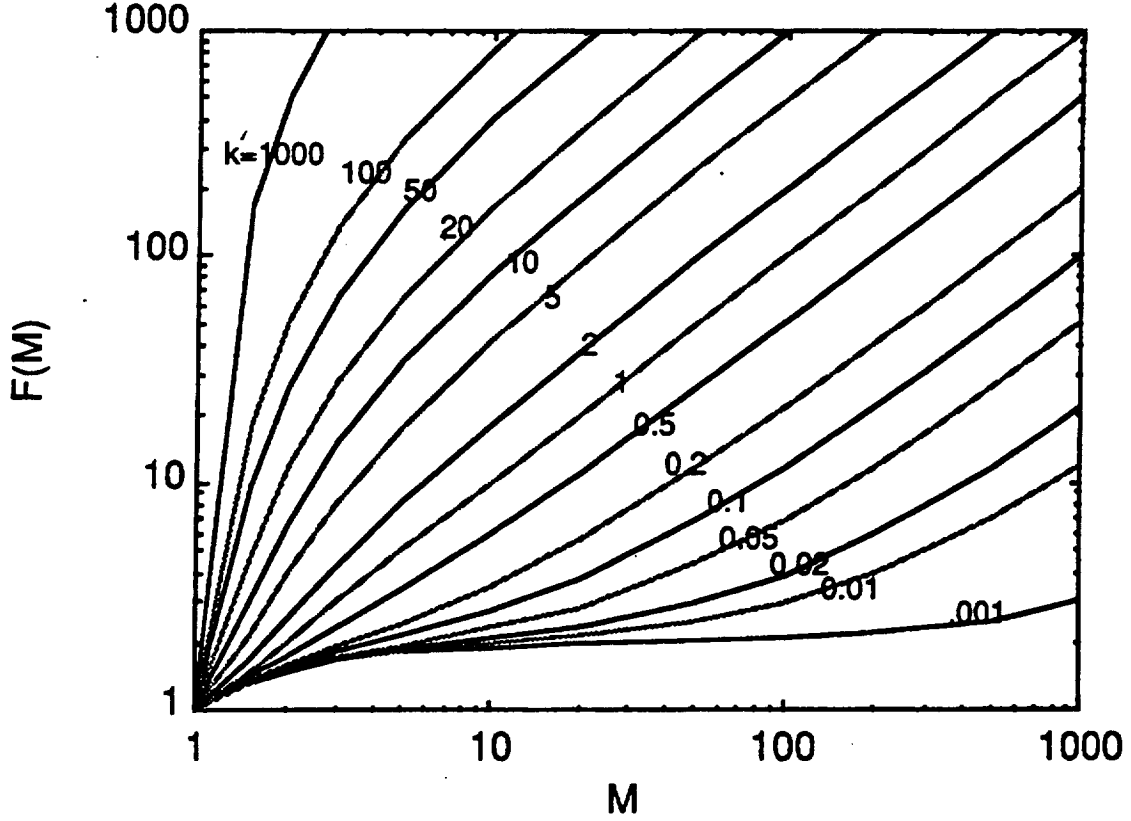


Figure 1-8: Excess noise factor $F(M)$ vs. M for constant k' , from McIntyre's equation.

Excess noise factor measurements were made with a HeNe laser using electron injection into the top p^+ layer. The experimental excess noise factor data is shown in Figure 1-9 where the dashed lines represent McIntyre's calculated theoretical curves. Figure 1-9(a) for the doped MQW APD clearly shows that for low gains ($M < 4$), the ionization ratio is greatly enhanced ($k=10-50$) as compared to that in bulk GaAs ($k=1.67$). This fact is clear evidence of the validity of our previous results for the gain values per stage which assume single carrier multiplication at low voltages. At higher voltages, however, the value of k is reduced since the holes gain more energy from the

applied electric field and are more likely to impact ionize¹⁴. The noise data for the conventional APD displayed in Figure 1-9(b) shows the high noise ($k \sim 1$) characteristics of the conventional APD even at low bias voltages. Excess noise factors at higher gain values were difficult to obtain since the dark current becomes large at high bias. Note that in Figure 1-9(a), the point where the excess noise data break away from the high k McIntyre curves corresponds to the breakdown voltage of the doped well APD. In addition, at high gains the k ratio for the doped well APD approaches the bulk GaAs value of 1.67. This is expected at high fields since the band bending resulting from the MQW structure becomes insignificant compared to that induced by the externally applied field.

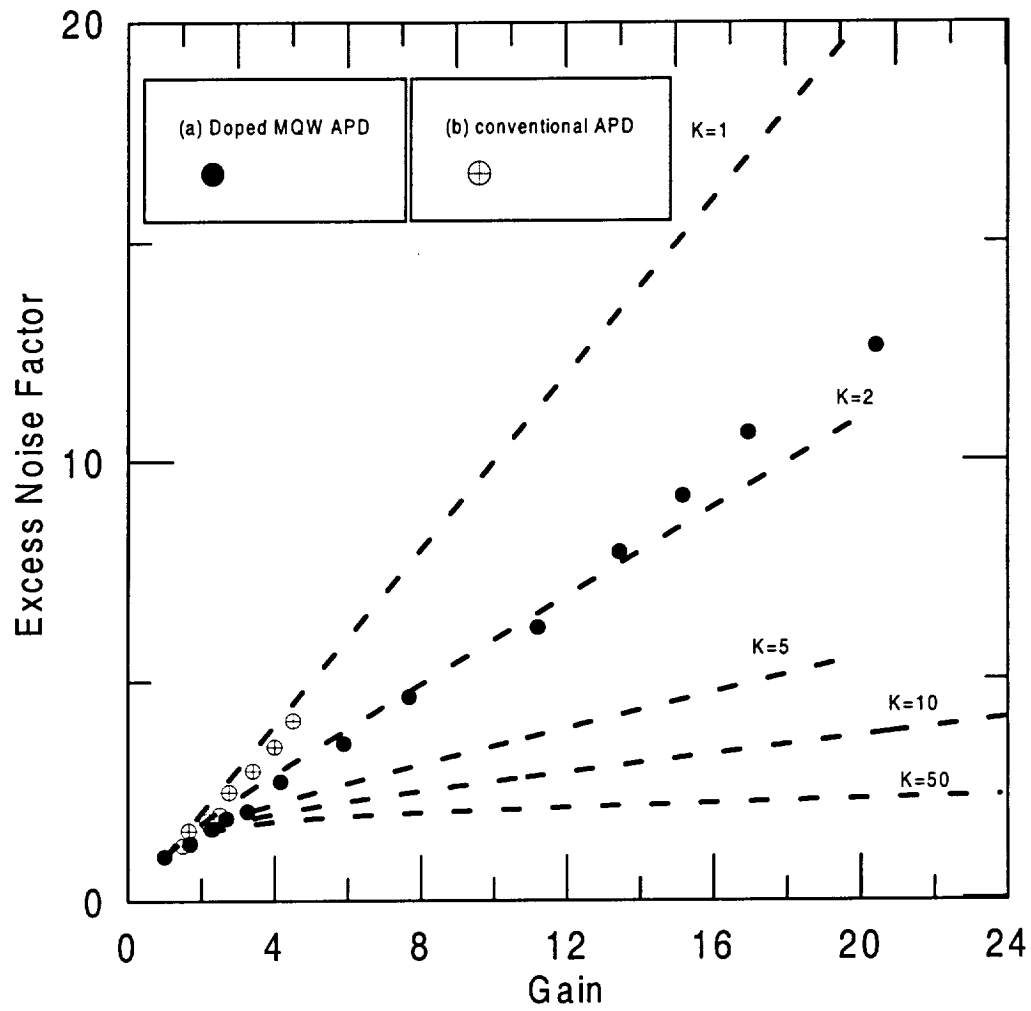


Figure 1-9: Excess noise factors for both (a) the doped MQW APD, and (b) the PIN APD. Dashed lines correspond to McIntyre theoretical curves for $k=1, 2, 5, 10$, and 50 .

2. Spectral Response Properties

2.1 APD Quantum Efficiency

The external quantum efficiency of a photodiode is defined as the number of electron-hole pairs generated at the output photocurrent per incident photon :

$$\eta_{ext} = (I_p / q) / (P_{opt} / h\nu) \quad [2-1]$$

where I_p is the photogenerated current resulting from the absorption of incident optical power P_{opt} at a given wavelength. Another related quantity is the responsivity which is defined as the ratio of the output photocurrent to the incident optical power:

$$\mathfrak{R} = \frac{I_p}{P_{opt}} = \frac{\eta q}{h\nu} = \frac{\eta \lambda (\mu m)}{1.24} \quad A / W \quad [2-2]$$

The quantum efficiency of a photodetector is primarily determined by the absorption coefficient α of the material. Figure 2-1 shows the measured intrinsic absorption coefficient for several materials used in photodetectors.¹⁵ From this figure, we can see that the room temperature absorption for GaAs material drops sharply around 0.9 μm . This long-wavelength cutoff wavelength is determined by the GaAs energy gap which is about 1.43 eV (~ 867 nm) at room temperature. At short wavelengths, the values of α become very large, and the radiation gets absorbed very rapidly near the surface where the recombination time is short. This will cause the photocarriers to recombine before they are collected by the junction region in a photodiode.

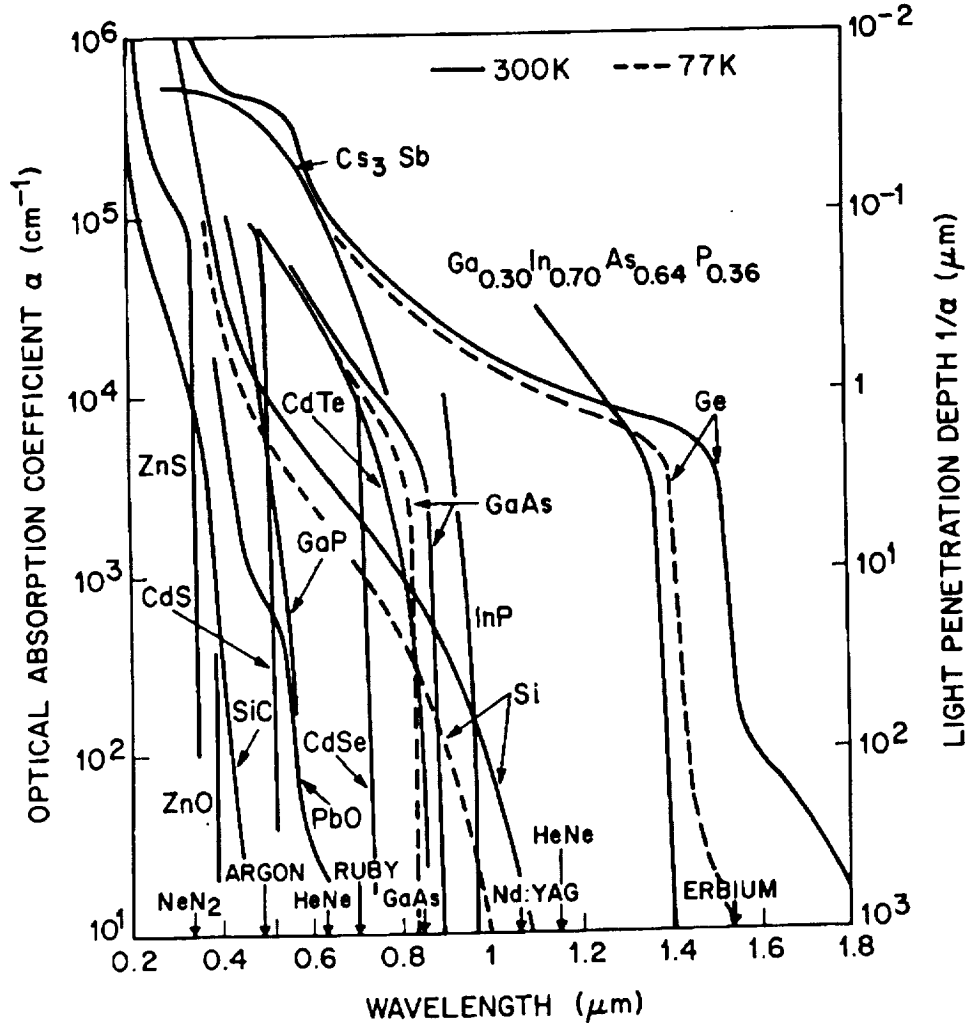


Figure 2-1: Optical absorption coefficients for different materials at 77 K and 300 K.¹⁶

The internal quantum efficiency of a photodiode η_{int} depends on the wavelength of the light as well as the thickness and doping of the absorption material. The absorption follows Beer's law and the internal quantum efficiency can be expressed as $\eta_{\text{int}} = 1 - \exp(-\alpha x_{\text{abs}})$, where α is the wavelength dependent absorption coefficient as shown in Figure 2-1, and x_{abs} is the thickness of the absorbing material. The "absorption length", l_a , is defined as $1/\alpha$ and gives the amount of material needed so that $1/e$ of the light would

be transmitted in the absence of reflections. The external quantum efficiency, η_{ext} , includes the effect of reflection as well as the various carrier recombination mechanisms.

2.2 Spectral Response Data

The long-wavelength behavior of the GaAs material was clearly demonstrated by spectral response measurements conducted on a doped PIN APD. The spectral response output is shown in Figure 2-3. The experimental curve was in good agreement with calculated spectral response data for similar devices. In order to calculate the experimental quantum efficiency, we measured the APD current output using a HeNe laser beam incident inside the p^+ ring with a total power of about 2 μW (inside a circular area with a 75 μm diameter). The experimentally calculated quantum efficiency at 632.8 nm was found to be about 19% for a doping level of $1 \times 10^{18} \text{ cm}^{-3}$. Table 2-1 shows the experimental external quantum efficiency as a function of the doping in the p^+ layer.

Table 2-1: Measured quantum efficiencies (at 633 nm) as a function of doping for a GaAs PIN¹⁷

N_A in cm^{-3}	η_{ext} in %
1.0×10^{18}	18.9
2.0×10^{18}	9.0
3.5×10^{18}	7.7

The simulated quantum efficiency curve (assuming zero reflection and no surface recombinations) is shown in Figure 2-3 for a 3 μm PIN photodiode at zero bias. The external source power density was maintained at 0.01 W/cm^2 . Using a 75 μm APD with an active area of $1.6 \times 10^{-4} \text{ cm}^2$, the total incident power on the top p surface is calculated to be about 1.6 μW . This is comparable to the HeNe laser power incident on the surface during the quantum efficiency experiment. Notice how the theoretical quantum efficiency at 633 nm is about 27% which is considerably higher than our experimental value. This is largely due to the loss of light due to surface reflection and to surface recombination mechanisms which were unaccounted for in our simulation. Even though reflection is neglected in the model, the maximum external quantum efficiency does not reach 100% due to the presence of various carrier recombination mechanisms (SRH, Auger, etc.) which were previously described.

It is possible to increase the quantum efficiency of the device through the introduction of a heavily doped p^{++} GaAs top layer which will help create a high-field region to enhance the diffusion of photogenerated electrons toward the depletion region. Figure 2-4 shows the calculated improvements in quantum efficiency for various doping differences between the 0.1 μm p^{++} layer and the 1 μm p^+ layer. According to the model, it

should be possible to increase the quantum efficiency by over a factor of 10 for the high energy part of the spectrum (0.2-0.4 μm). In addition, the response throughout the visible spectral region becomes more uniform as can be seen in Figure 2-4.

Note that the introduction of the thin layer does not make a significant difference to the quantum efficiency when the top absorption region is heavily doped ($3 \times 10^{18} \text{ cm}^{-3}$). This is due to increased carrier recombination in the highly doped 1 μm region which tends to reduce the number of carriers diffusing toward the depletion region.

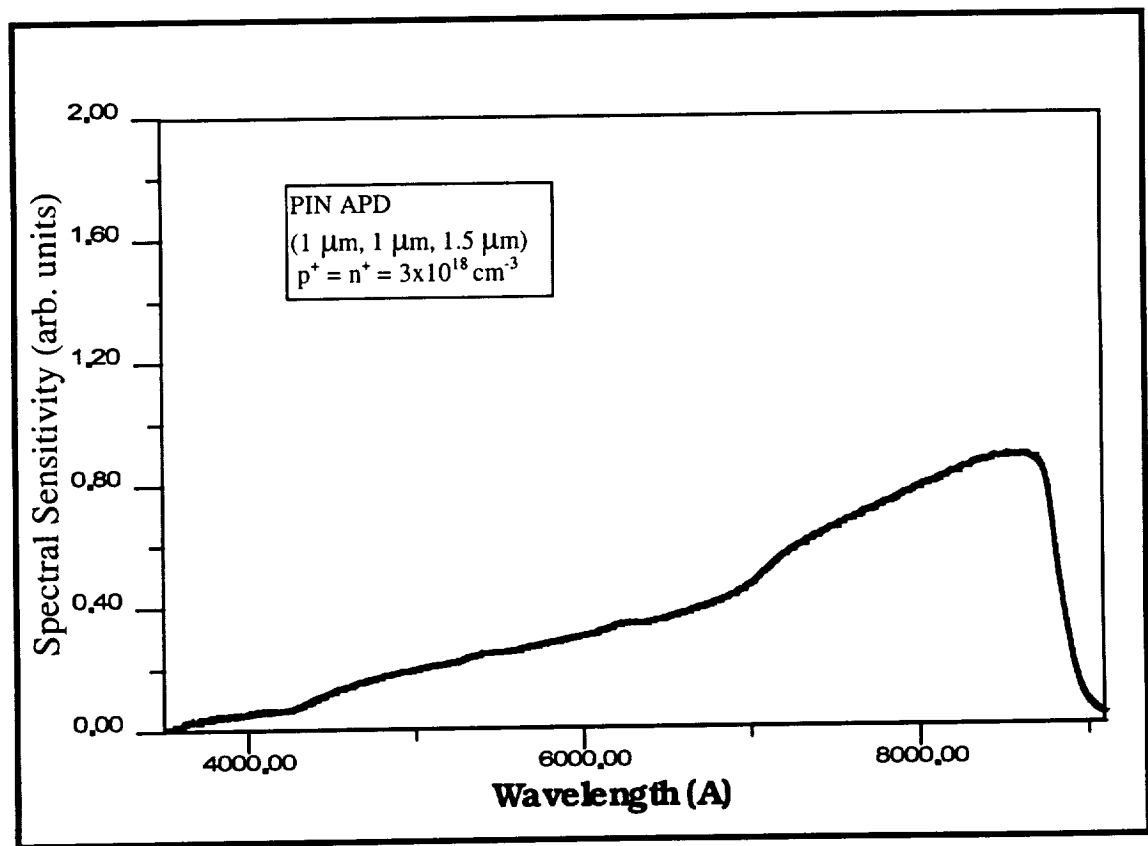


Figure 2-2: Spectral response measurement of an MBE grown PIN APD.

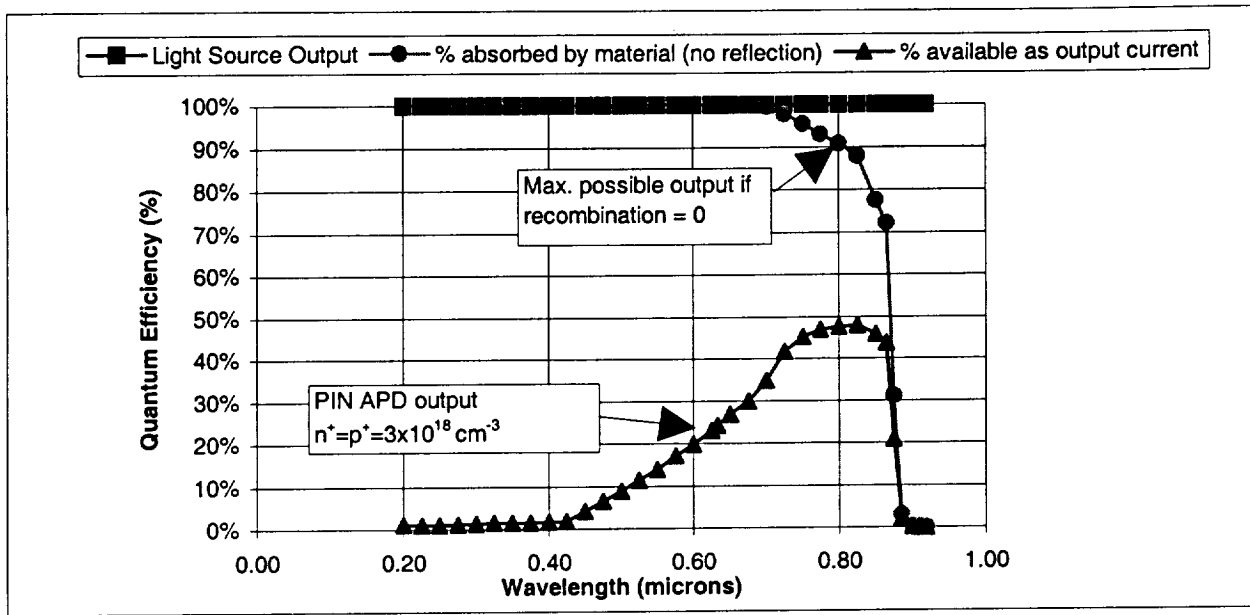


Figure 2-3: Theoretical quantum efficiency plot (zero reflection) for a 3 μm GaAs PIN (1 μm /1 μm /1 μm) APD where $p^+ = n^+ = 3 \times 10^{18} \text{ cm}^{-3}$.

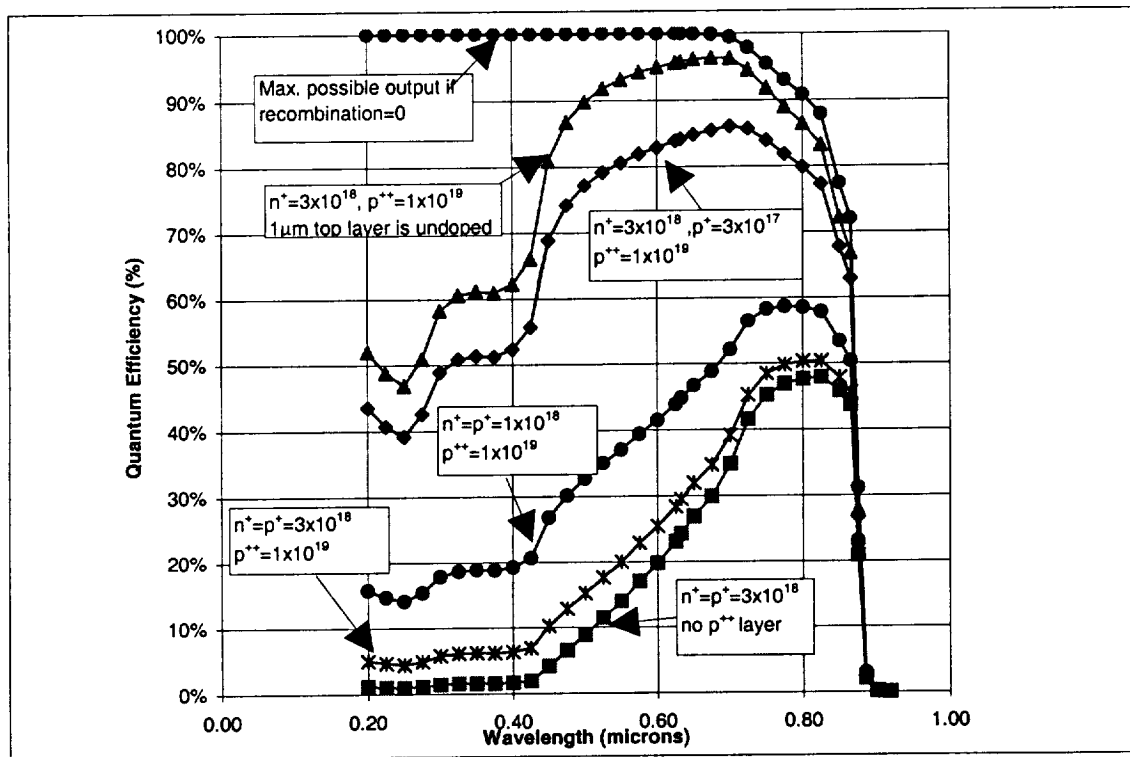


Figure 2-4: Theoretical plots showing possible improvement in quantum efficiency for a PIN APD through variations in the doping concentrations and the introduction of a thin (0.02 μm) top p^{++} layer (all doping concentrations given are in cm^{-3})

3. Time Response Characteristics

The requirements of high bit rate (> 4000 Mb/s) lightwave communication and image processing systems have necessitated the development of fast photodiodes that have higher sensitivity than the PIN detector. Improvement in sensitivity while maintaining wide bandwidths can only be provided using an amplification mechanism within the photodetector itself as is the case in avalanche photodiodes. For optimum operation, an APD must meet the following criteria¹⁸: (1) the electric field in the multiplication region must be high enough to produce sufficient gain; (2) the electric field in the absorbing region must be low enough so that the tunneling component of the dark current is negligible; (3) the depletion region must extend far enough into the absorbing region so that diffusion effects are negligibly small. These requirements impose rather severe constraints on the doping concentrations and thickness of the epitaxial layers of the device. Transient response measurements can provide valuable information on the speed performance and the various factors that affect the bandwidth of APDs. Since carrier diffusion plays a major role in determining the time response of a photodetector, one would expect the speed to depend greatly on the depletion characteristics of the APD. The more depleted a structure is, the shorter the distance the carriers will have to diffuse, and the faster the response time will be. This will be shortly demonstrated with experimental time response data. But first, in order to better understand the results that were obtained, a brief discussion the various physical effects that limit the frequency response of a photodiode will be presented.

3.1 Bandwidth Limitations of Photodetectors

The bandwidth of a Separate Absorption and Multiplication Regions (SAM) APD is determined by five physical effects:

1. *Transit Time*: This is the time it takes the generated carriers to travel through the depleted region under the effect of the electric field. There are two types of transit times in an APD. The primary carrier transit time corresponding to the photogenerated carriers, and the secondary carrier transit time required for the multiplied carriers of opposite type to retrace the steps of the primary ones. Transit times for electrons and holes (τ_e and τ_h) are usually calculated using the ratio of the distance traveled and the "saturation" velocity of the appropriate carrier.

2. *Carrier diffusion time*: In the undepleted regions of the device, carrier transport must take place by diffusion rather than drift. Because of the absence of electric field in the absorption layer of a PIN APD, the photogenerated carriers must diffuse in order to reach the avalanche region. This results in the slowing of the device's response. An oscilloscope trace of the transient output of such a device would show both a "fast" and a "slow" component. The fast component is due to carrier drift, and the slow one, referred to as the "diffusion tail", is due to diffusion from the undepleted regions of the device.

3. *RC time constant*: There is a fundamental limit on bandwidth due to the capacitive transient charging effects which arise from the depletion region capacitance of the device and the combined resistance R of the load and the device.

4. *Hole trapping*: In heterojunction APDs, there is a possibility of carrier delay caused by traps present at the heterojunction interface. This effect is related to the abruptness of the heterojunction, the barrier height, the temperature, and the effective mass of the carrier. Because the effective mass of holes is larger by an order of magnitude than that of electrons, trapping is more likely to occur for holes than electrons. This phenomena is known as "hole trapping" and it can be minimized through the use of graded composition layers instead of abrupt heterojunctions.

5. *Avalanche buildup time*: For single carrier ionization, one only needs to consider the transit time through the multiplication layer. For dual carrier ionization, however, there is a feedback process that introduces a time delay through the multiplication region. This is called the avalanche buildup time. In an APD, there is a buildup time τ_{aval} associated with the avalanche gain process which tends to limit the time response of the photodetector. The primary avalanche build-up time for electron initiated multiplication is :

$$\tau_{aval\ n} = \frac{K_{disp}}{v_n + v_p} \int_0^w \exp\left[-\int_0^x (\alpha - \beta) dx'\right] dx \quad [3-1]$$

where K_{disp} is a correction factor; v_n and v_p are the electron and hole velocities. The physical origin of K_{disp} is the electron/hole displacement current which arises from the space-charge induced E-field resulting from the motion of carriers.¹⁹ The closer the value of α is to that of β , the more secondary carriers are generated, and the higher the avalanche build-up time as can be seen from equation [3-1].

3.2 Overall Photodetector Bandwidth

In the case of a PIN photodetector where absorption takes place in the junction, the basic limitations to the response time are due to the RC and the transit times of the primary carriers. The overall PIN time constant is usually approximated by the square root of the sum of squares of the RC and transit time constants:

$$\tau_{\text{pin}}^2 = \text{Max}(\tau_h, \tau_e)^2 + \tau_{\text{RC}}^2 \quad [3-2]$$

As was previously mentioned, in an APD, there are two different transit times arising from the primary carriers traveling to, and secondary carriers traveling from, the multiplication region. In addition, there is the avalanche buildup time (proportional to gain) which is proportional to the multiplication process. The actual APD frequency response is a complicated function of all of these processes. Hollenhorst²⁰ and Roy²¹ have developed complicated transfer functions and matrix expressions to estimate the time constants for arbitrary structures. For approximation purposes, the RC time constant is usually treated as being non-correlated with the rest of the time constants. In addition, the primary hole transit, hole trapping, avalanche buildup and secondary electron transit events can be assumed to occur in series, one following the other. In this case, the sum of squares can be used to approximate the total time constant as well.

In the following section, we will present some of the experimental data and attempt to provide the proper interpretations as they relate to the processes described above.

3.3 Experimental Results

Figure 3-1 shows the pulse response for an unbiased doped MQW device with a 2.5 μm MQW region. The APD was mounted on a 50 GHz Tektronix sampling scope and

was excited with an 810 nm 50 ps laser pulse. As can be seen from the figure, the unbiased APD output pulse has a rise time of about 317 ps, a fall time of 2.5 ns and a full width at half max (FWHM) of about 1.4 ns. The oscilloscope trace is a convolution of the 50 ps gaussian laser pulse with the output response of the APD. The fast rise time component typically follows the relaxation oscillation of the laser pulse. The falling edge of the pulse shows a "fast" and a "slow" component. The slow component at the trailing edge is usually attributed to either charge trapping at interface states or diffusion of carriers in the undepleted regions of the structure.²² Diffusion will limit the speed of the device as long as there are undepleted regions in the structure and a separate absorption layer is being used. At high bias, trapping is no longer an issue, and the device response is limited by the transit time and the RC time constant. In the following, it will be demonstrated that such a slow response is due largely to diffusion effects in the partially depleted APD structure.

Figure 3-2 shows the response of the above APD under bias (low gain). The fall time and the FWHM have now dropped to 819 and 952 ps, respectively, corresponding to about a 32% increase in the speed of the device. If the bias is increased further, as shown in Figure 3-3, those values drop to 570 and 593 ps, respectively, with a speed increase of about 58%. The large dependence of the pulse's width and tail on the applied bias is a clear indication of a diffusion-limited time response.

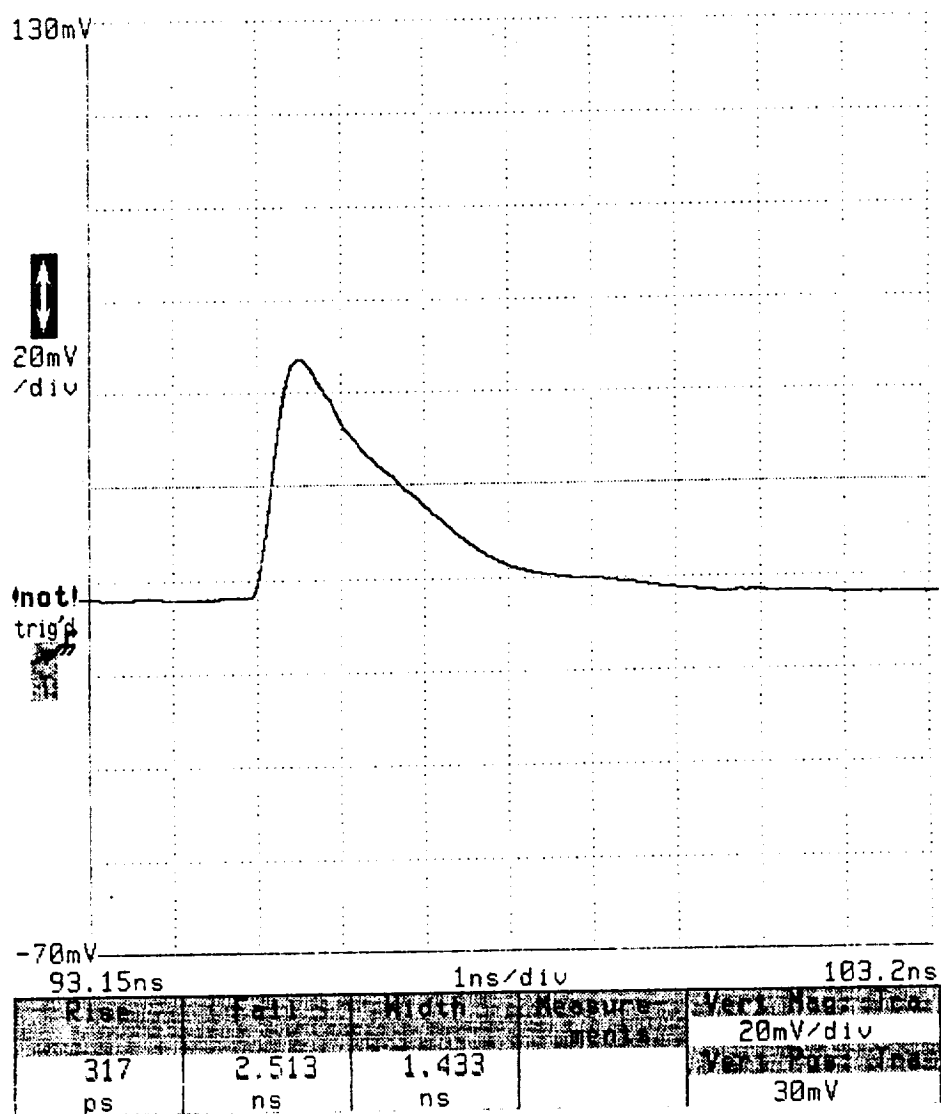


Figure 3-1: Oscilloscope trace of the pulse response of a doped 2.5 μm MQW APD under no bias

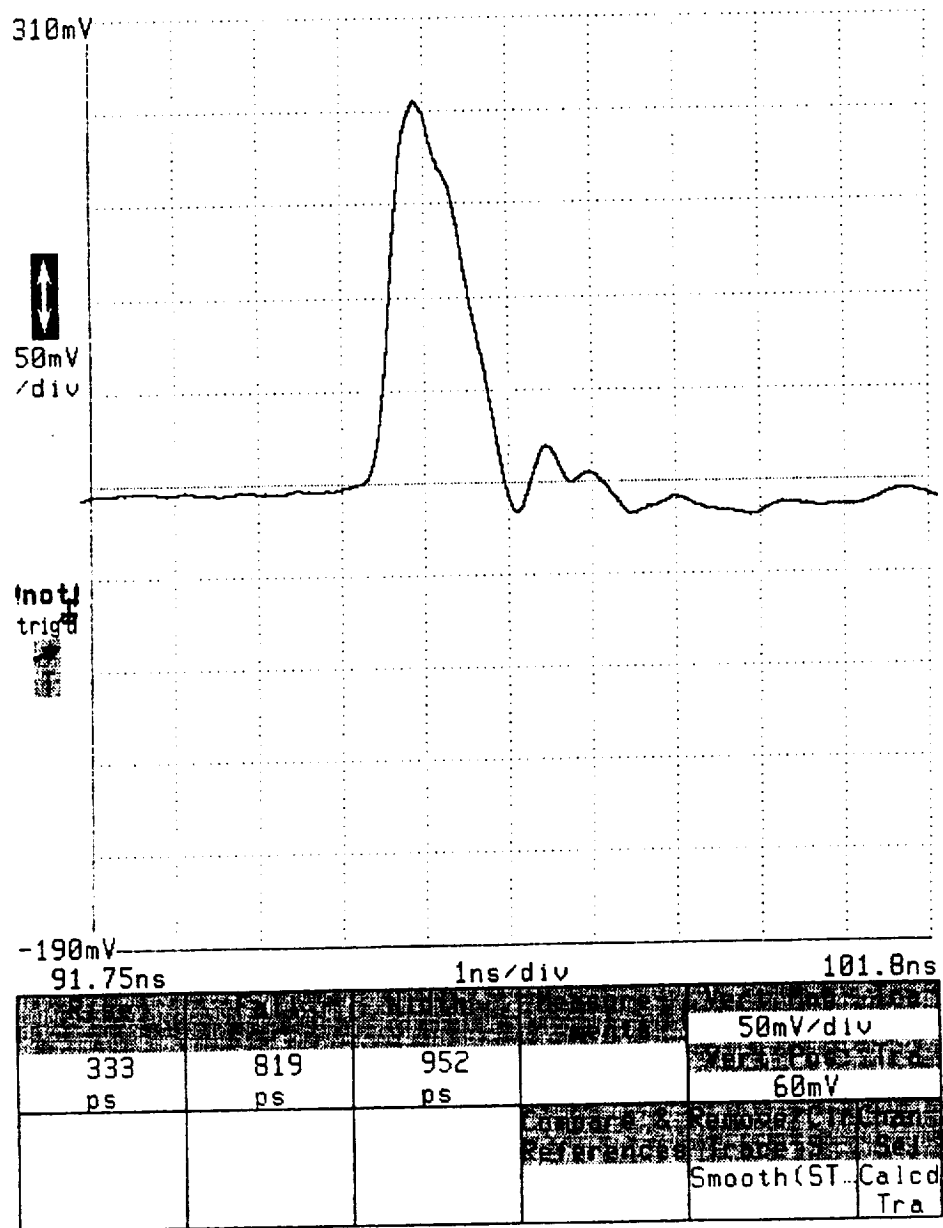


Figure 3-2: Oscilloscope trace of the pulse response of a doped 2.5 μm MQW APD under low bias

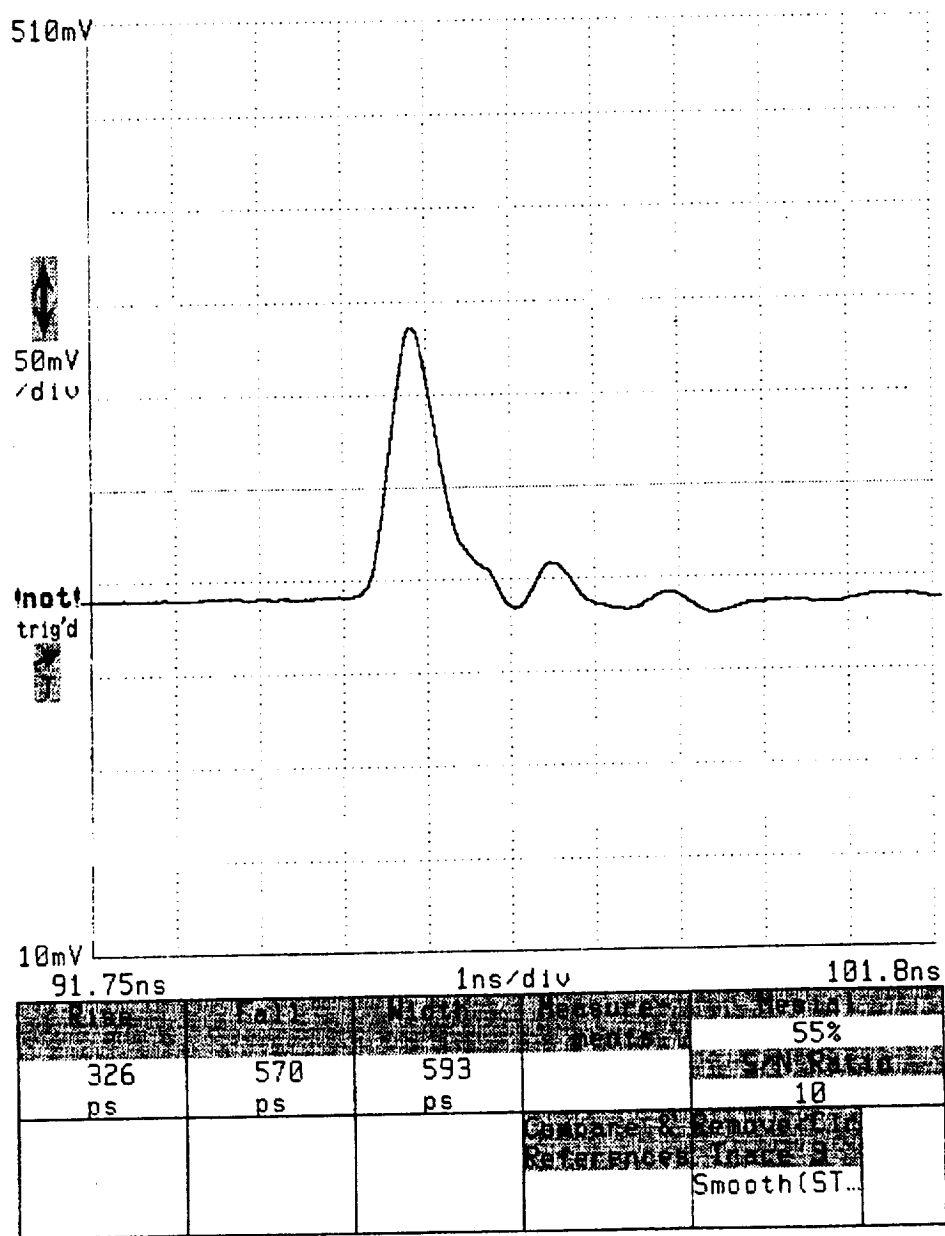


Figure 3-3: Oscilloscope trace of the pulse response of a doped 2.5 μm MQW APD near breakdown.

The diffusion-limited effect can be verified further by examining the depletion width characteristics of the structure obtained from CV measurements. This is shown Figure 3-4 where the zero bias depletion width of the MQW and PIN structures are about 0.1 μm and 1.4 μm respectively. These values can be used to estimate the drift time constants from $\tau_{\text{drift}} = W_{\text{depleted}} / v_{\text{sat}}$ where W_{depleted} is the width of the depleted region, and v_{sat} is the saturation velocity taken to be in the order of 10^7 cm/s. The drift time constants for the MQW and PIN structures were estimated to be 1 ps and 14 ps, respectively. Similarly, the ratio of the diffusion time constants between the MQW and the PIN structures is directly related to the ratio of the undepleted widths ($\tau_{\text{diff(MQW)}} / \tau_{\text{diff(PIN)}} = W_{\text{undep(MQW)}} / W_{\text{undep(PIN)}} = 2.2$). Using that ratio and the sums of squares approximation, the diffusion time constants for the MQW and PIN structures were calculated to be 1.5 ns and 0.7 ns, respectively. This result shows that the MQW structure is largely diffusion-limited due to the presence of a large undepleted region. This behavior is largely due to a mismatch in the doping balance between the n and p doping layers in the MQW structure. On the other hand, the PIN APD shows a much faster time response (Figure 3-5) due to the fact that the structure is largely depleted even at zero applied bias Figure 3-4. Therefore, the time response limitations for this structure are mainly due to 1) diffusion time in the top p^+ and bottom n^+ layers, and 2) transit time in the intrinsic field region of the structure. Diffusion time in the cap layers can be optimized by varying the thicknesses as well as the doping concentrations. The transit time can be shortened by increasing the field (applied bias) across the junction. This, however, begins to create an additional delay near breakdown due to the increase in the avalanche buildup time.

The "ringing effect" seen in the oscilloscope trace following the output pulse was due to the impedance mismatch between the APD circuit and that of the oscilloscope sampling head. With applied bias, the impedance of the APD changes due to the increased conductivity of the structure. Note that the relative magnitude of the pulses in the case of the MQW APD is not representative of the gain of the device since a variable resistor was used in an attempt to match circuit resistance and thus limited the voltage applied at the oscilloscope.

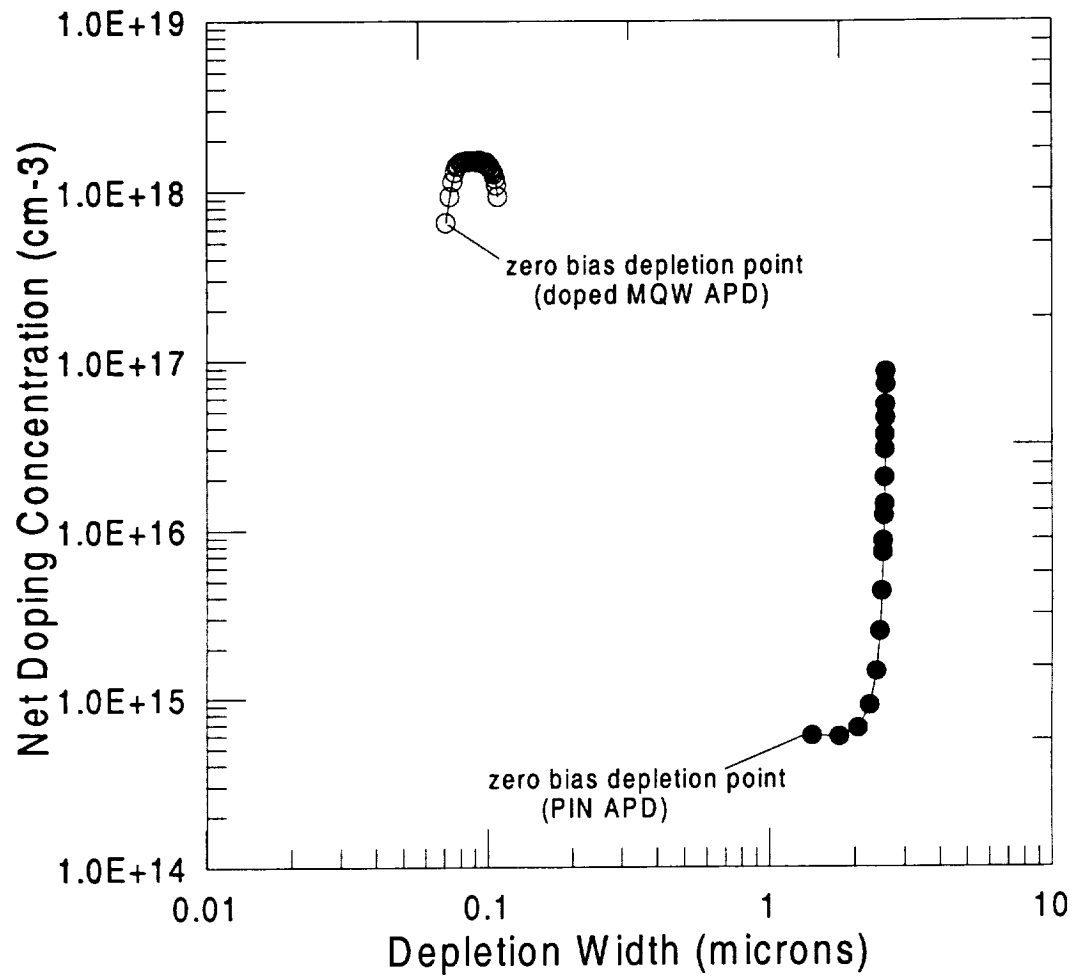


Figure 3-4: Depletion width plot showing the increase in the depletion region with applied bias for a largely undepleted doped (2.5 μm) MQW APD.

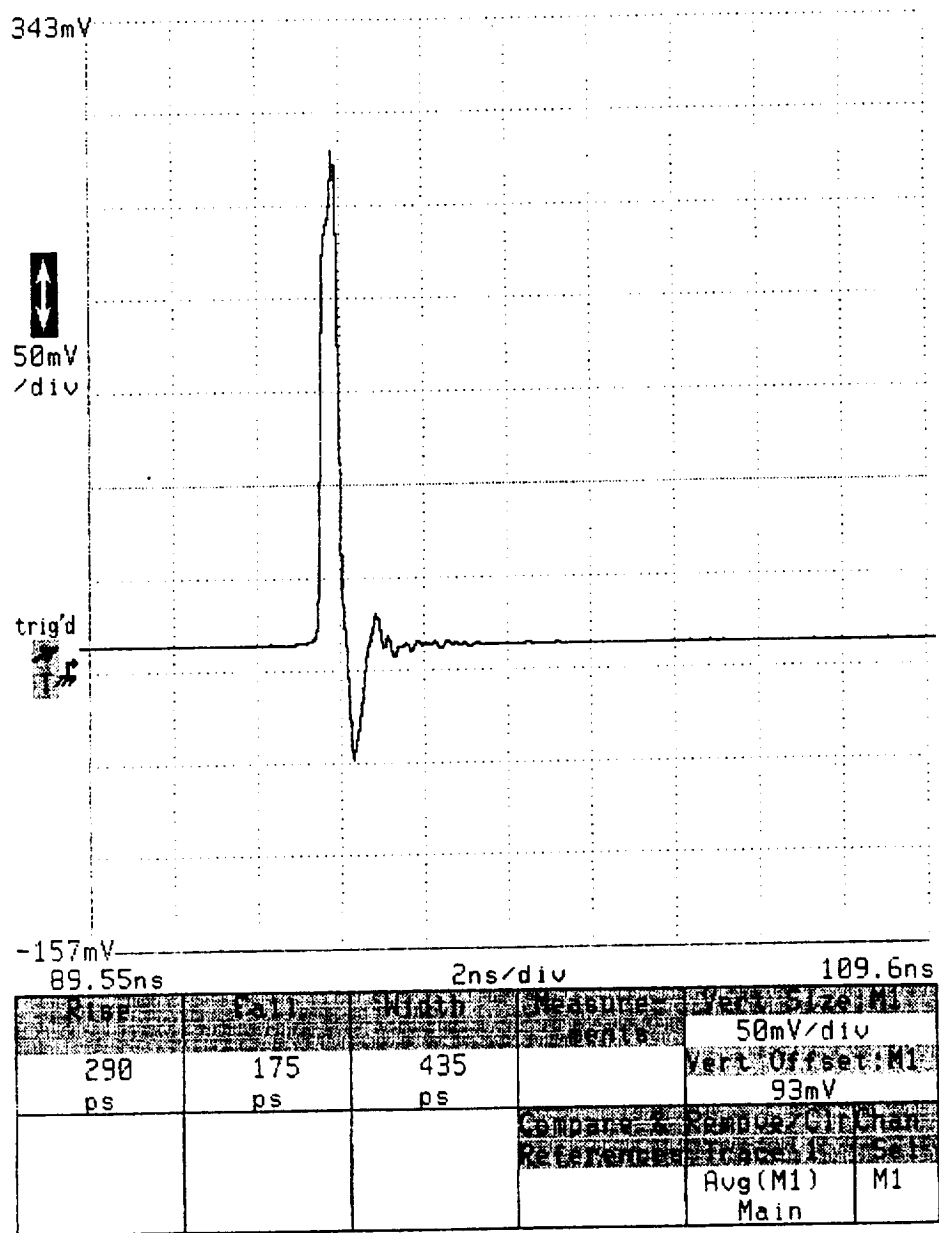


Figure 3-5: Oscilloscope trace of the pulse response of an undoped 2.5 μm PIN APD with no applied bias.

4. Temperature Dependence

4.1 Impact Ionization vs. Tunneling

In most materials, the mechanisms of junction breakdown can be attributed to both impact ionization and tunneling. One way of determining the dominant process is by examining the temperature dependence of the junction's IV characteristics. Since the energy bandgap of GaAs decreases with increasing temperature, one would expect the breakdown voltage due to the tunneling effect to have a negative temperature coefficient leading to a decrease in breakdown voltage with increasing temperature as shown in Figure 4-1. This is because a smaller applied field would be needed to reach the same current levels at higher temperatures. On the other hand, avalanche breakdown has a positive temperature coefficient where the breakdown voltage increases with increasing temperature due to the shorter mean free path of carriers at higher temperatures. Increasing T increases the phonon vibrations of the lattice, thereby increasing the probability of premature scattering and reducing the ionization coefficients for a given E -field. Other sources of leakage current, such as generation-recombination and diffusion also tend to increase with increasing T . The result of such an effect is shown in Figure 4-2 by the experimental gain data obtained from the IV curves of a doped MQW APD. According to Tyagi²³, the breakdown voltage is related to temperature through the following linear relationship:

$$V_B(T) = V_B(T_0)(1 + \beta(T - T_0)) \quad [4-1]$$

where $\beta > 0$ in junctions where impact ionization dominates. Such a linear dependence was shown experimentally to be valid by Forrest et al.²⁴ in the case of p-n junctions.

However, our experimental measurements have shown that for doped MQW junctions, the data can be best fit using a third degree polynomial as shown in Figure 4-3.

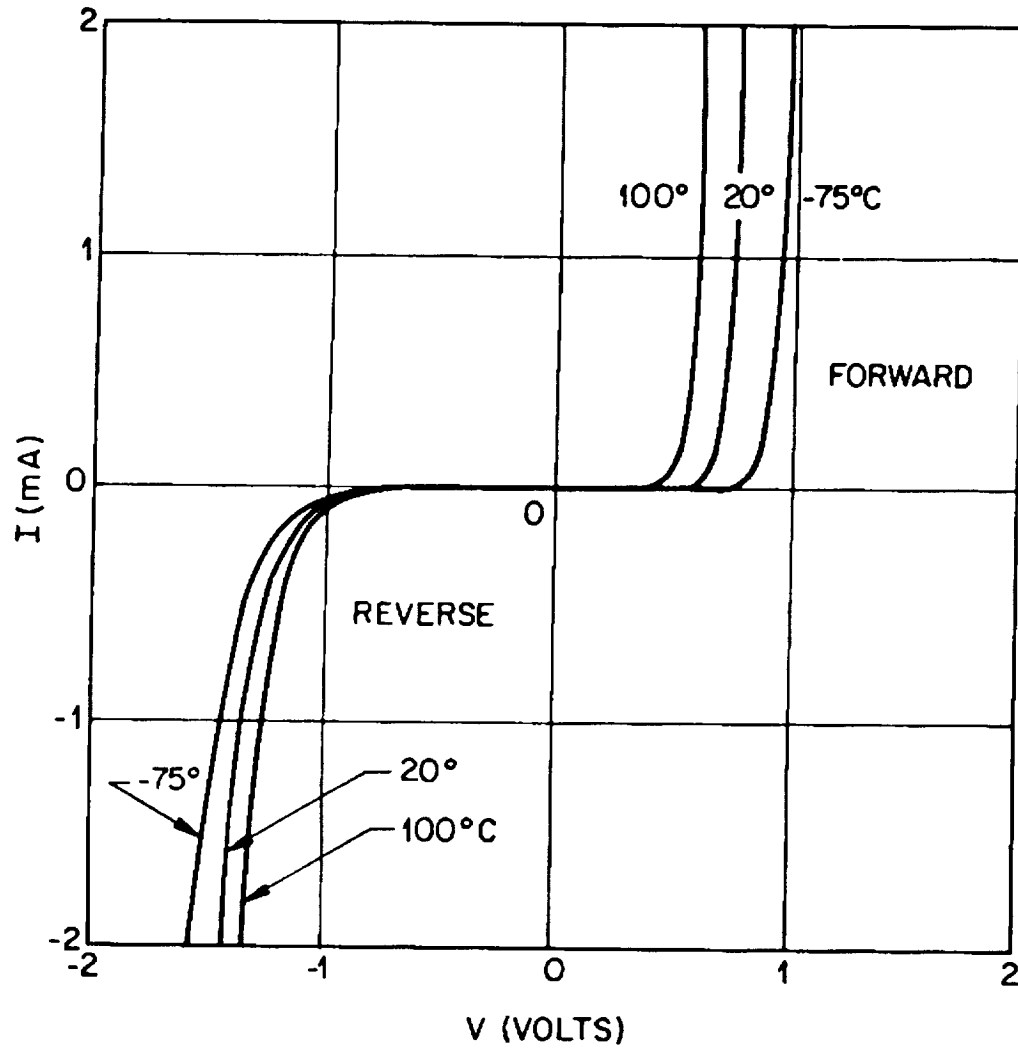


Figure 4-1: IV characteristics of tunneling breakdown²⁵.

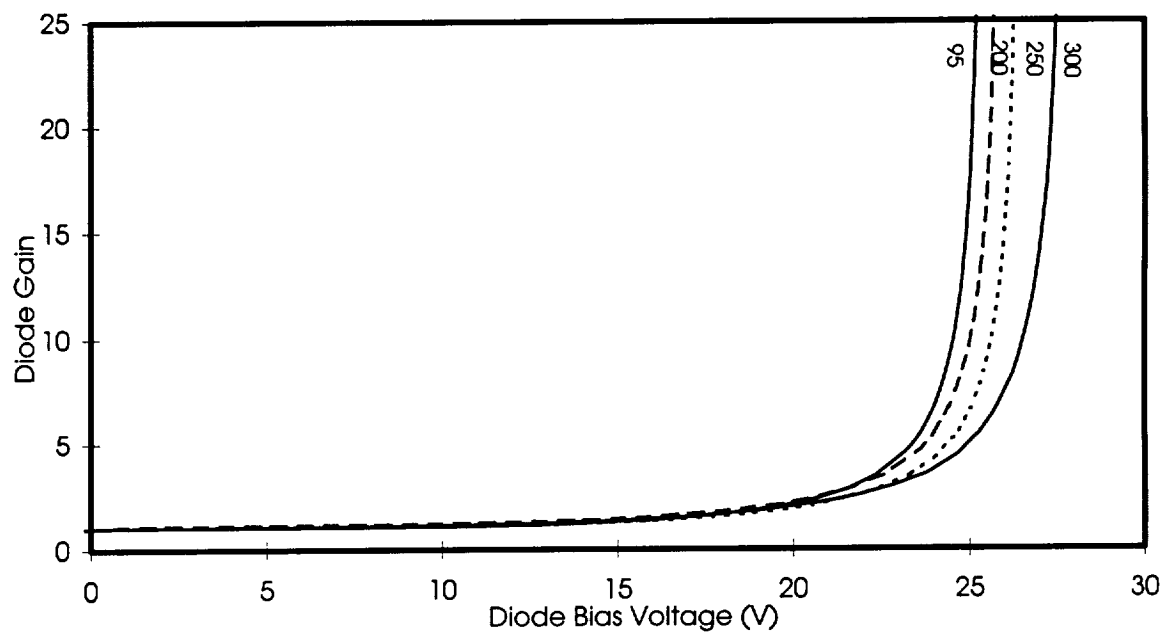


Figure 4-2: Measured gain as a function of temperature (in K) for a doped MQW APD.

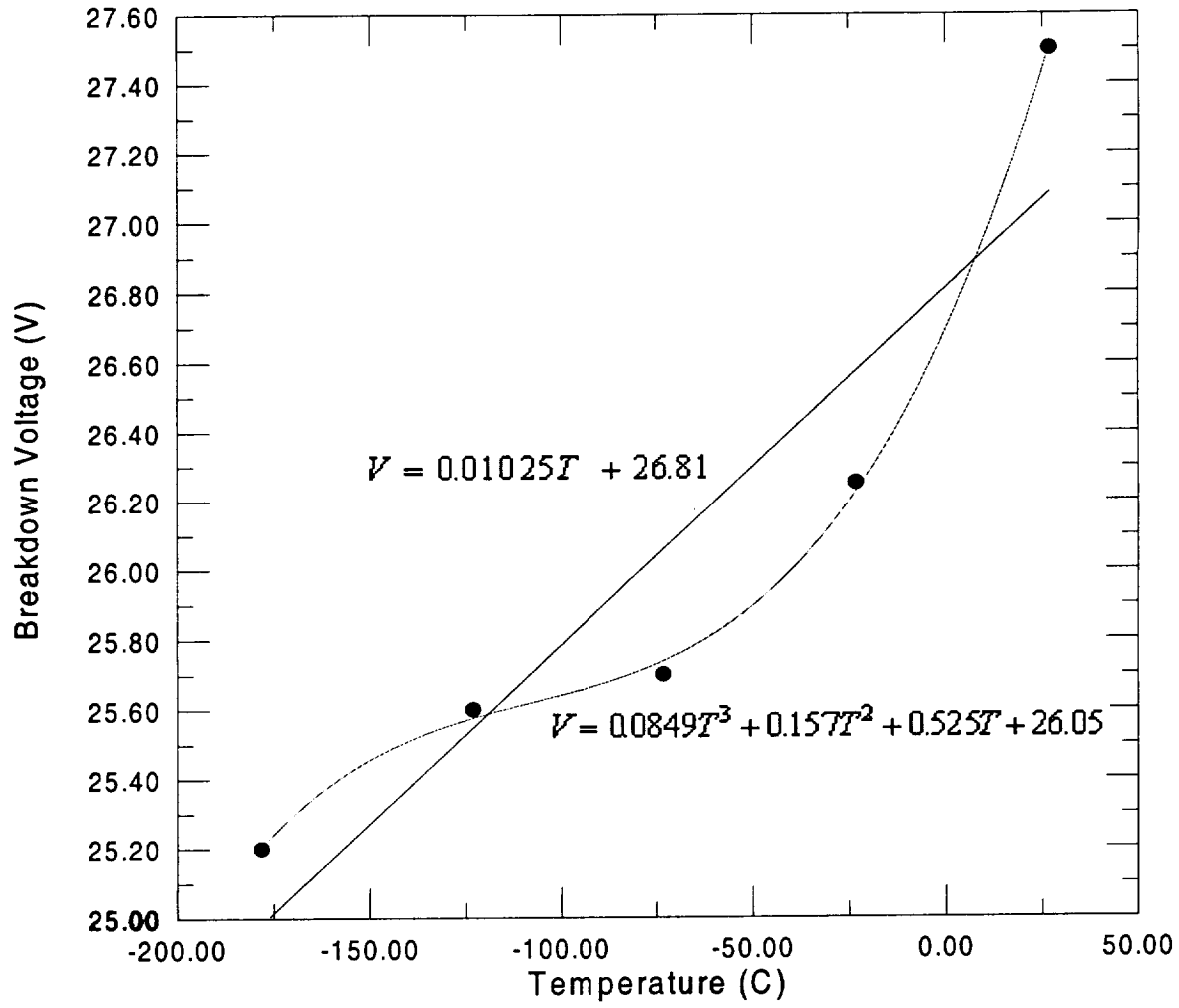


Figure 4-3: Experimental data and fits for a doped-well MQW APD.

4.2 Junction Capacitance and Temperature

The PN junction capacitance is given by:

$$C_j = \frac{K_s \epsilon_0 A}{\left[\frac{2K_s \epsilon_0 (V_{bi} - V_A)}{q} \frac{(N_A^- + N_D^+)}{N_A^- N_D^+} \right]^{1/2}} \quad [4-2]$$

which shows that C_j is directly proportional to $[N_A^- N_D^+ / (N_A^- + N_D^+)]^{1/2}$. The number of ionized donors and acceptors are given by:¹⁶

$$N_D^+ = N_D \left[1 - \frac{1}{1 + \frac{1}{g_D} \exp\left(\frac{E_D - E_F}{kT}\right)} \right] \quad [4-3]$$

$$N_A^- = \frac{N_A}{1 + g_A \exp\left(\frac{E_A - E_F}{kT}\right)} \quad [4-4]$$

where g_D is the ground state degeneracy of the donor impurity level and is equal to 2 since a donor level can accept one electron with either spin or can have no electron. On the other hand, g_A is the ground-state degeneracy factor for acceptor levels and is equal to 4. This is because in GaAs as well as in Ge and Si, each acceptor impurity level can accept one hole of either spin and the impurity level is doubly degenerate as a result of the two degenerate valence bands at $k=0$.

Therefore, by examining equations [4-2]-[4-4], it is clear that the capacitance is expected to decrease exponentially with decreasing temperature following the decrease in the ionized donor and acceptor densities. This behavior is illustrated in the experimental CV data in Figure 4-4 where the capacitance at low bias decreases exponentially toward a limiting value of about 2.8 pF in the case of a doped-well MQW structure. As the reverse bias is increased at a given temperature, the depletion width increases causing the capacitance to drop toward 2.8 pF corresponding to the capacitance value for maximum depletion of the structure.

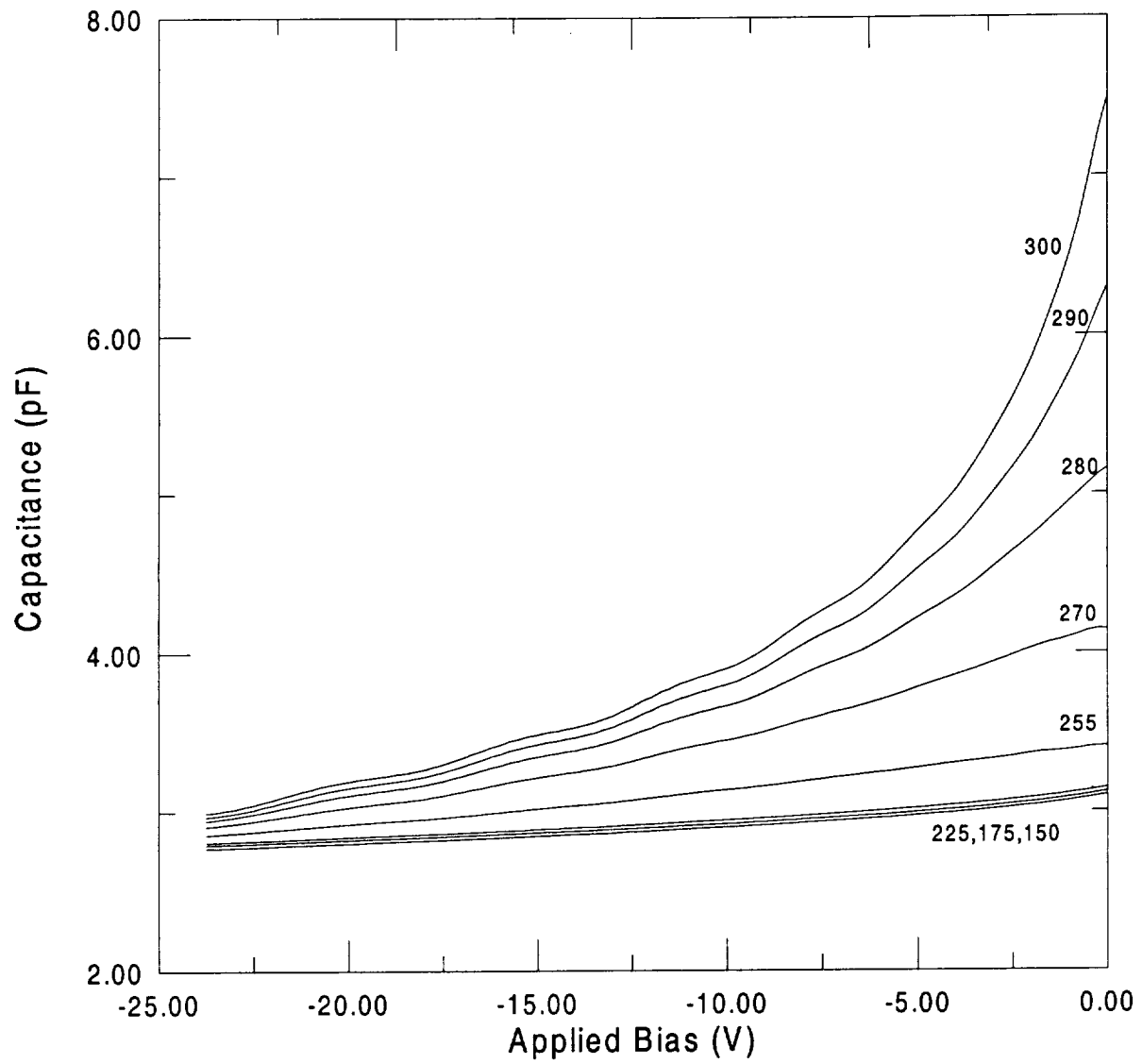


Figure 4-4: Experimental CV curves as a function of temperature for a doped-well MQW APD.

5. Effect of Variations in the Doping Profiles

The purpose of the following analysis is to use both theoretical and experimental evidence to determine the impact of doping imbalance and symmetry on the physical and electrical characteristics of doped MQW APDs. Theoretical models have been developed to calculate the electric field, valence and conduction band profiles, CV profiles, as well as carrier concentration versus depth profiles. Our models showed a strong correlation between the p- and n-doping balance inside the GaAs wells and the number of depleted stages and breakdown voltage of the APD. A periodic doping imbalance in the wells has been shown to result in a gradual increase (or decrease) in the electric field profile throughout the device which gave rise to partially depleted devices at low bias. The MQW APD structures that were modeled consisted of the standard structure with a 1 μm doped-well MQW region. These simulation results showed that in an APD with nine doped wells, and where the 50 \AA p-doped layer is off by 10% compared to the n-doped layer ($p=1.65 \times 10^{18} \text{ cm}^{-3}$, $n=1.5 \times 10^{18} \text{ cm}^{-3}$), half the stages were shown to be undepleted at low bias which was a result of a reduction in the E-field near the p^+ cap layer by over 50% from its value in the balanced structure. Experimental CV and IV data on similar MBE grown MQW structures have shown very similar depletion and breakdown characteristics. The models have enabled a better interpretation of the experimental data and relate some of the observed peculiarities in the IV and CV curves directly to the doping profile in the MQW structure.

5.1 Introduction

Several characterization techniques have been devised to analyze and understand the optical and electrical properties of APDs²⁶. These include various experiments such as IV measurements which are used to determine the gain properties of the device, and CV measurements which are used to calculate carrier concentration versus depletion width profiles. However, the data obtained using such experimental techniques are not always easy to interpret and relate back to the physical processes taking place inside the structures. In addition, the practical limitations inherent in the growth and fabrication of large quantities of devices with different structural designs add even more complexity to the problem because of the large number of variables involved in the process.

In what follows, a more practical approach will be presented to analyzing the experimental data obtained using IV and CV experiments and specifically those relating the doping profile characteristics to device properties. Accurate theoretical models of MQW APD structures have been developed using AtlasII, Silvaco's two-dimensional device simulation framework. These models were used to provide graphical representations of the spatial variations of the electric field across the biased structure, as well as conduction and valence band diagrams of the GaAs/AlGaAs MQW structure before and after breakdown. In addition, avalanche breakdown simulations and small signal ac analysis were used to extract IV and CV curves in order to compare the data from the models to those obtained directly from our experimental devices. Both electron- and hole-injected photocurrent solutions were obtained by simulating a 632.8 nm monochromatic light source with spot power of about 1 W/cm² incident on the devices'

front and back surfaces. CV solutions were obtained using small signal analysis at a frequency of 1 MHz and with a signal magnitude of 0.03 V. All our analysis were conducted using Newton's two-carrier method²⁶ and the generation rate of electron-hole pairs due to impact ionization was modeled according to Selberherr²⁷.

5.2 Theoretical Results

The APD structure used in this model consisted of a top and bottom p⁺ and n⁺ doped ($3 \times 10^{18} \text{ cm}^{-3}$) GaAs layers with thicknesses of 1 μm . The middle region was made up of 10 periods of alternating layers of GaAs (500 Å) and Al_{0.42}Ga_{0.58}As (500 Å). The GaAs wells were similarly doped with p-i-n layers whose thicknesses and doping concentrations were treated as variable parameters for the purpose of our study. When a reverse bias is applied, the combined effect of the applied electric field, the built-in field, and the conduction band offset enhances the ionization process of electrons in the GaAs. The holes, on the other hand, are subjected to a smaller valence band discontinuity and therefore gain less energy than the electrons.

In this study, the widths of the p and n doping layers were held constant at 50 Å, and that of the intrinsic layer at 100 Å. The doping imbalance ($\frac{|p|-|n|}{|n|}$) was varied between zero and 100 percent. Figure 5-1 shows the corresponding CV plots obtained for similar devices with 0%, 33.3%, and 100% doping imbalance. As expected, the capacitance of the device is lowest when the p and n doping are perfectly matched since the net carrier concentration throughout the device is reduced to zero.

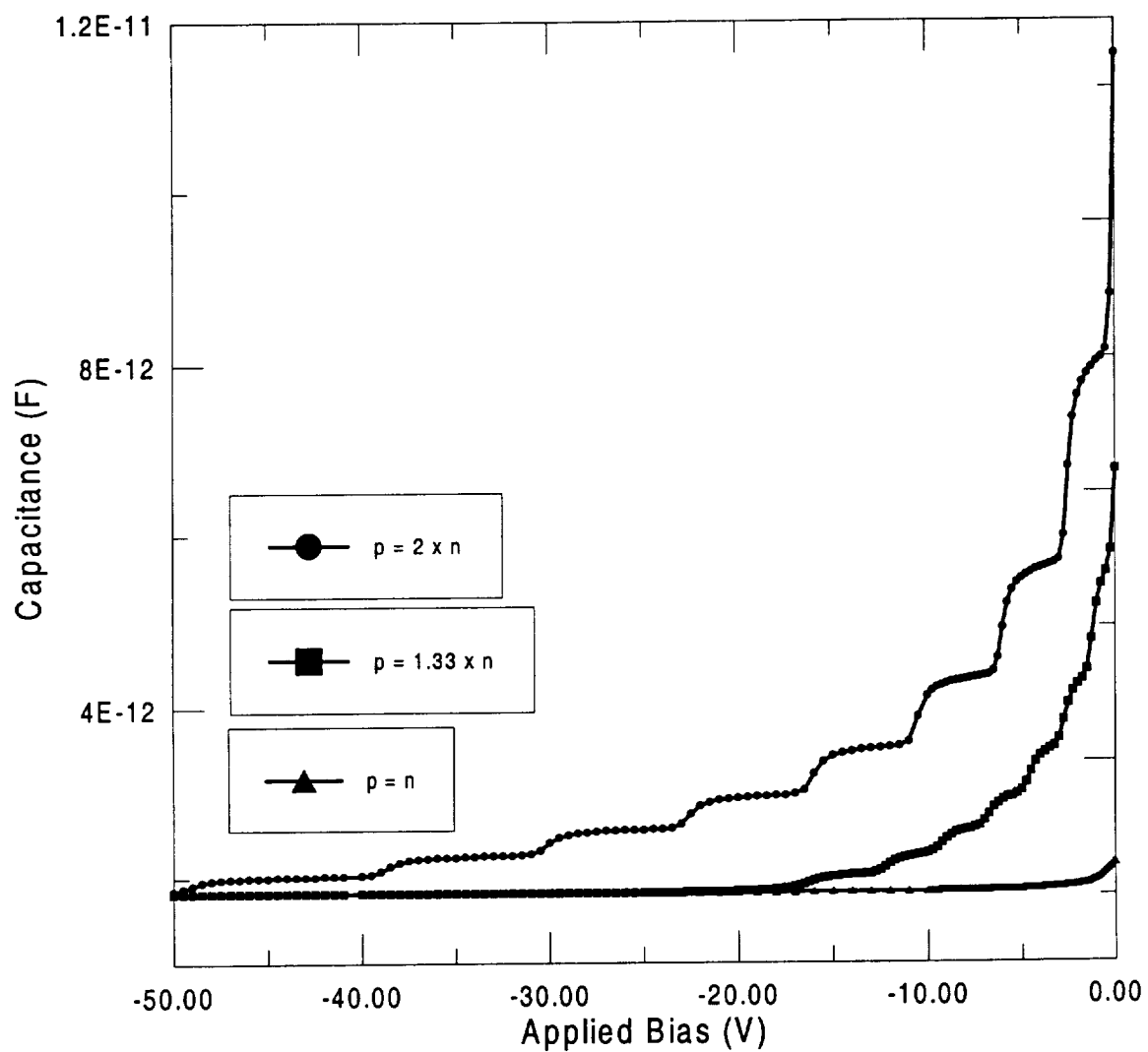


Figure 5-1: Comparison of theoretical CV data obtained for the same APD MQW structure where the doping mismatch in the wells was varied between 0% and 100%.

However, the avalanche breakdown voltage as depicted by the IV curves in Figure 5-2, seems to be highest when p is equal to n . This is due to the fact that a doping mismatch would result in a gradual increase of the electric field throughout the device which would cause impact ionization to take place at a lower bias point. Therefore, a large doping imbalance would actually lower the bias at which breakdown occurs. This, however, comes at the expense of a large undepleted region which could limit the quantum efficiency and severely hurt the time response characteristics of the photodiode.

Figure 5-4 shows the calculated carrier concentration versus depletion width profile for similar APD structures with 0%, 10%, 20%, 33.3%, and 100% doping mismatch. In the case where $p=n=1.5 \times 10^{18}$, it can be seen that the device is fully depleted at zero bias. The number of undepleted stages begins to increase when increasing the offset between p - and n -doping. In the case where the p -doping is twice that of n , only about 20% of the device is depleted at zero bias. In order to better understand the effect of the doping imbalance on the MQW structures, it is helpful to examine the valence and conduction band diagrams Figure 5-5, as well as the electric field spatial profile shown in Figure 5-3. Devices corresponding to 0%, 33%, and 100% doping imbalances are modeled at a reverse bias of 20 V. As is seen from the two figures, the electric field is uniformly symmetric, and the MQW region is equally depleted in the case where $p=n$. However, as p gradually increases, the electric field becomes progressively lower near the top p -layer which results in non-uniform depletion of the MQW structure. The effect of such non-uniform depletion on the device's photocurrent can be clearly seen in the IV plots shown in Figure 5-2. In the case of electron injection, the photocurrent at zero bias is about four

orders of magnitude lower than that for hole injection. Gradually, the electron-injected photocurrent increases as the device is depleted until it reaches about 6×10^{-8} A corresponding to that of the hole-injected photocurrent. Therefore, a doping imbalance where $p > n$ can greatly reduce the device's external quantum efficiency in the case of electron injection. Such an effect is not as pronounced in the case where $n > p$ due to the smaller valence band discontinuity faced by the injected holes.

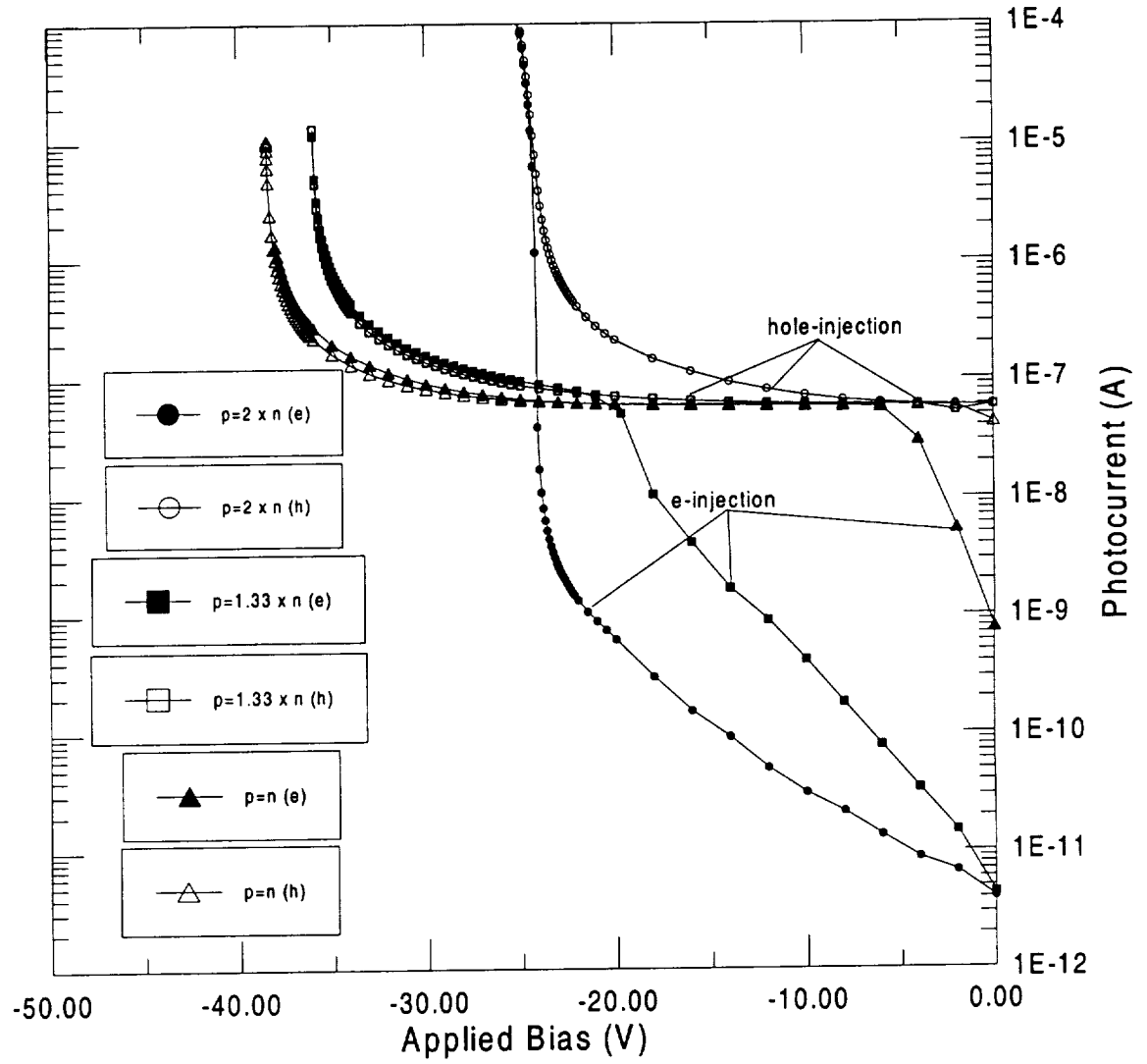


Figure 5-2: Comparison of theoretical light IV data obtained for the same APD MQW structure where the doping mismatch in the wells was varied between 0% and 100%

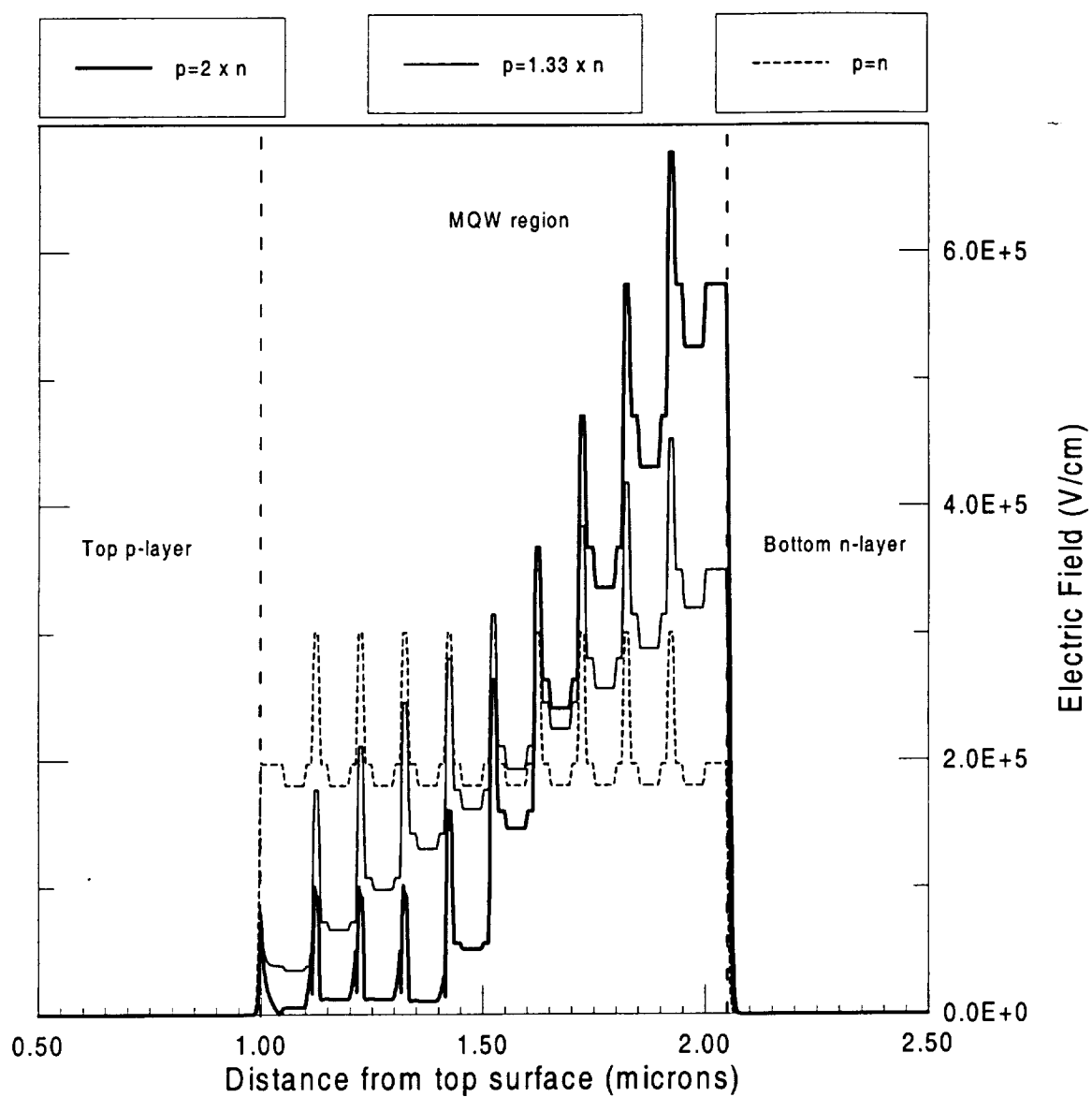


Figure 5-3: Theoretical electric field profiles of the same MQW APD structures with different doping imbalance in the wells ($V = -20$ V).

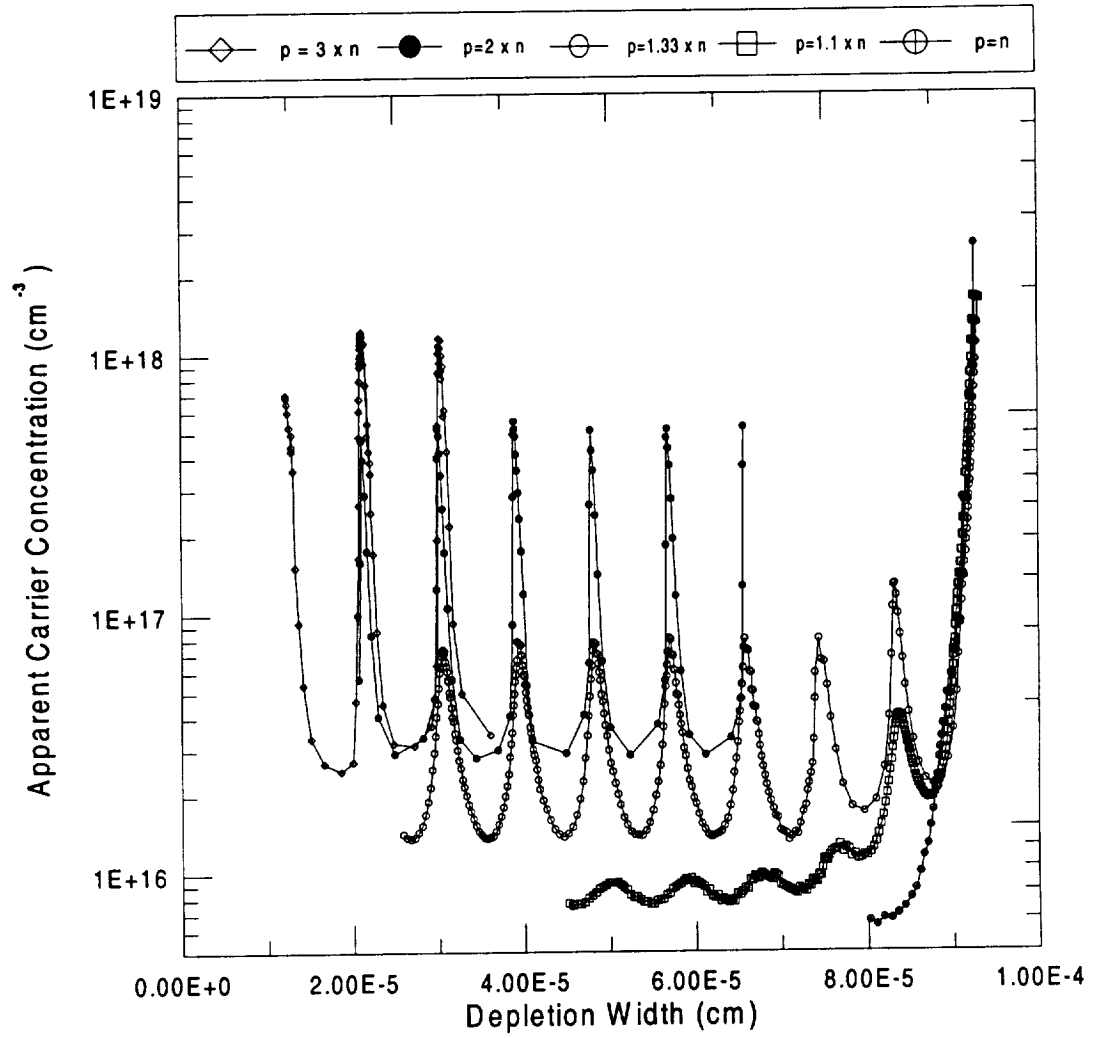


Figure 5-4: Calculated doping profiles versus depletion width using theoretical CV data obtained from AC analysis on similar MQW APDs with different doping imbalances in wells.

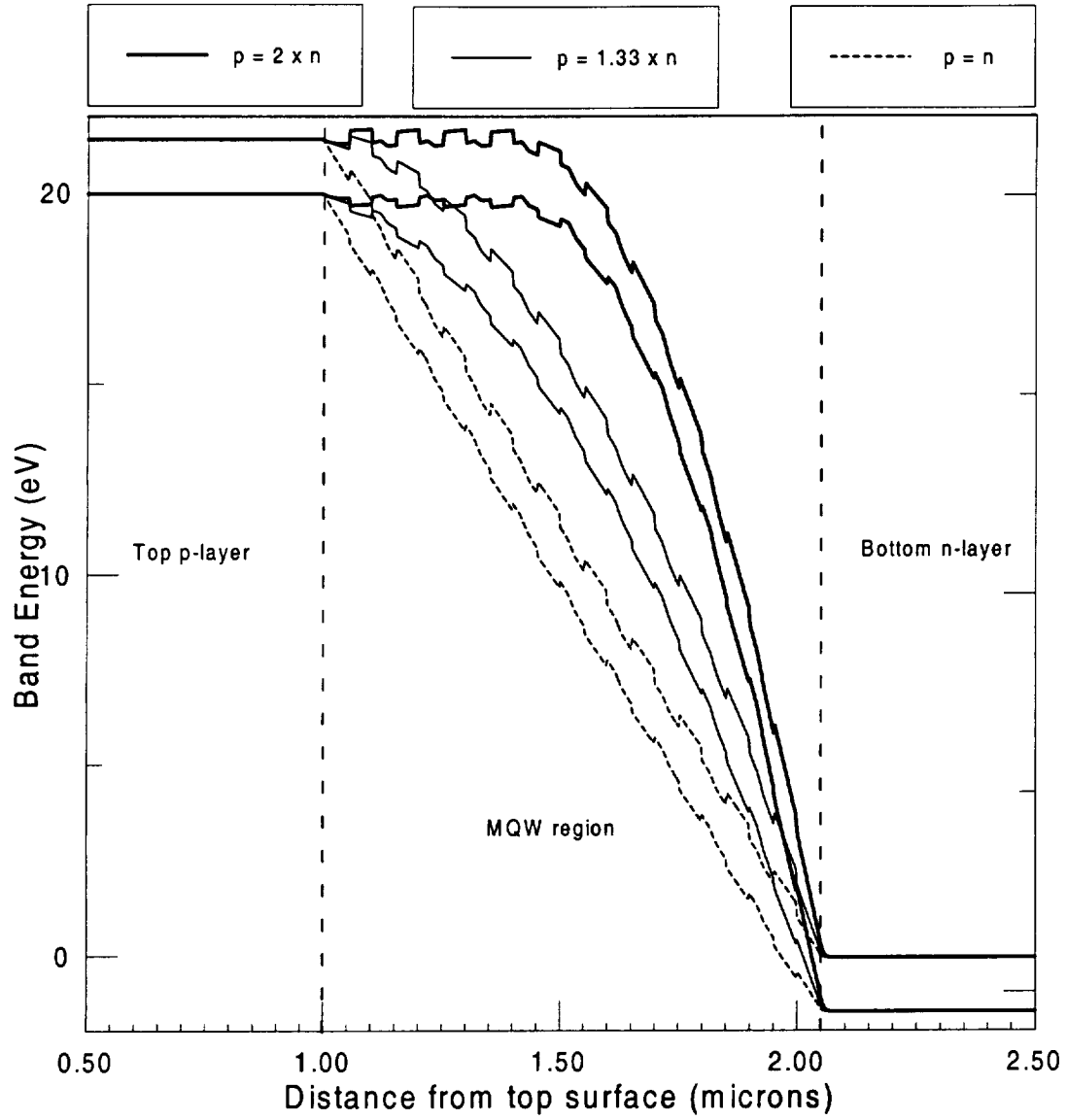


Figure 5-5: Theoretical band diagrams of the same MQW APD structures with different doping imbalance in the wells ($V = -20$ V)

5.3 Experimental Results

CV measurements were performed on all devices at 1 MHz. The CV data was then analyzed to calculate the depletion widths and carrier profiles for the structures. Figure 5-6 shows the net carrier concentration plots for four doped-well MQW devices labeled as APD1 through APD4. All devices have similar geometries except for the p- and n-doping in the wells which was varied between 0.5×10^{18} and $1.5 \times 10^{18} \text{ cm}^{-3}$. It is interesting to see that even though APD1 and APD2 were expected to have very similar properties, their CV and carrier concentration profiles were quite different. APD1 was almost fully depleted at zero bias, while APD2 was not and only reached full depletion right before breakdown. Note that the peak positions in the carrier profile of APD2 do not quite agree with the 1000 Å period in the MQW structure and with the doping profiles obtained using our models. This discrepancy is due to the fact that carrier concentration profiles calculated from the experimental CV data assumed a one sided depletion²⁸ which apparently does not hold true for the experimental devices. Other sources of error in the experimental data result from the inability to accurately account for parasitic capacitance between the devices and the metal contacts and bonding wires in the measurement system. In addition, note how the average net carrier concentration in the experimental doping profile gradually increases up to the top GaAs well where it then drops indicating that the doping imbalance is not the same throughout the structure.

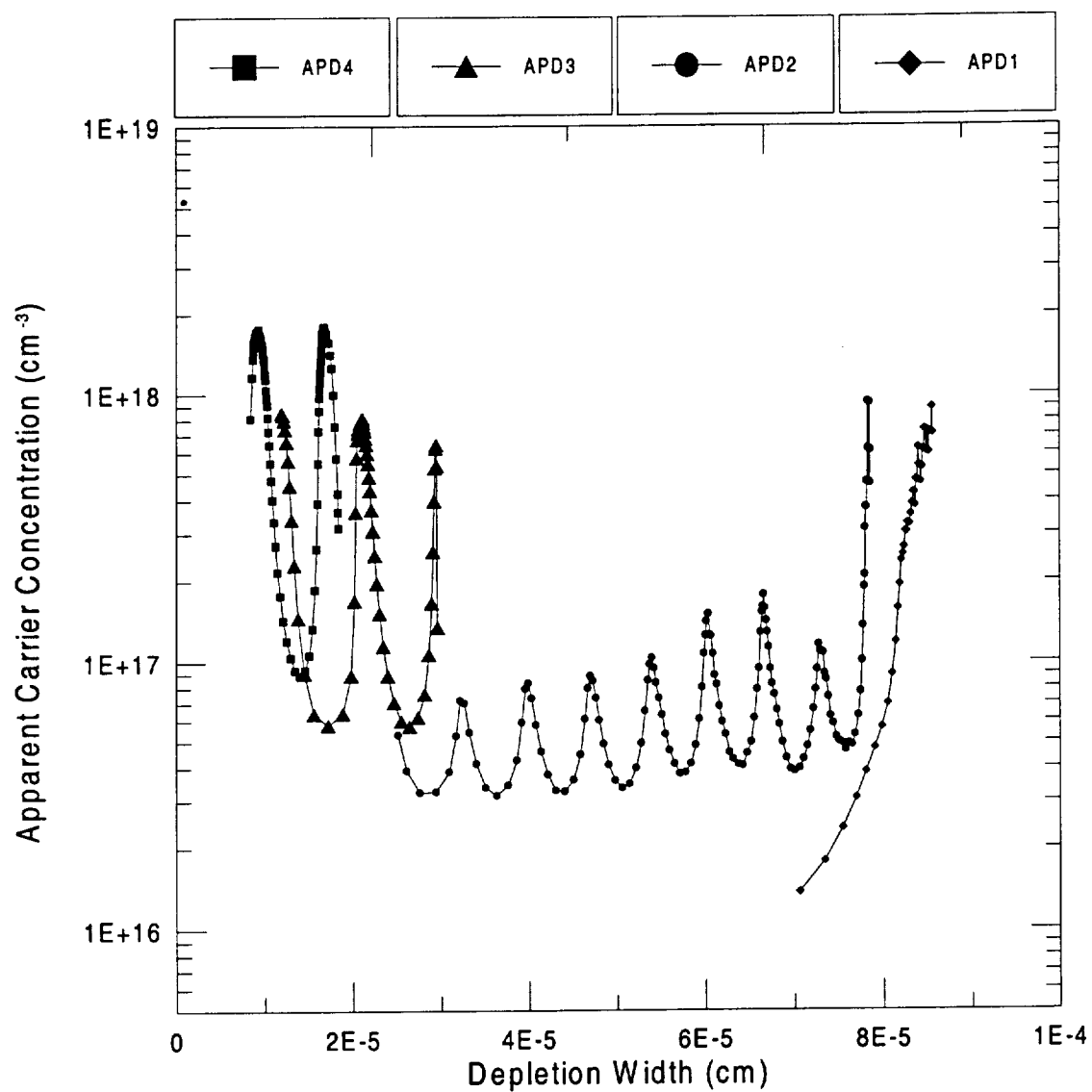


Figure 5-6: Calculated doping profile versus depletion width using experimental CV data from four 10-period doped-well MQW APDs.

The spatial resolution of the CV measurements (both experimental and theoretical) was limited by the Debye length which is about 40 Å at room temperature for a doping level of $n=1.5 \times 10^{18} \text{ cm}^{-3}$. Since the thickness of the doped layers in the wells was of the same order of magnitude (50 Å), abrupt changes in the doping concentration could not be accurately measured. Therefore, it is generally difficult to relate the apparent carrier concentration obtained from the experimental devices to the actual doping imbalance in the wells. However, using our theoretical carrier profile where the actual doping imbalance is fully known, it is possible to estimate the actual doping mismatch in every doped layer in the experimental MQW device by superimposing both the experimental and theoretical data. Therefore, we can roughly conclude that the average doping imbalance in APD1 is far less than 10% which resulted in full depletion at zero bias, while that in APD2 is between 30% and 40% where full depletion of the MQWs was achieved near breakdown around 27 V. In the case of APD3 and APD4, the situation was quite different. Apparently, the p- and n-doping mismatch was so large (~ 200 %) to the point where only partial depletion of two wells was accomplished before avalanche breakdown. By examining the electron injected photocurrent curves, we can easily conclude that for both APD1 and APD2, the doping mismatch is such that $n > p$, while in APD3 and APD4, the situation is reversed. This can be clearly understood by comparing the light IV data to the theoretical curves. The gradual increase in the electron injected photocurrent in APD3 and APD4 is an indication of trapping of injected electrons by the AlGaAs barriers near the p-layer where the device is undepleted. Therefore, according to the models shown in Figure 5-4, the average p-doping in the wells must be larger than that

of n . In the case of APD1 and APD2, no trapping seems to take place since the low bias photocurrent is much higher and relatively flat. Therefore, these two devices have undepleted regions near the n -layer which indicates that $n > p$.

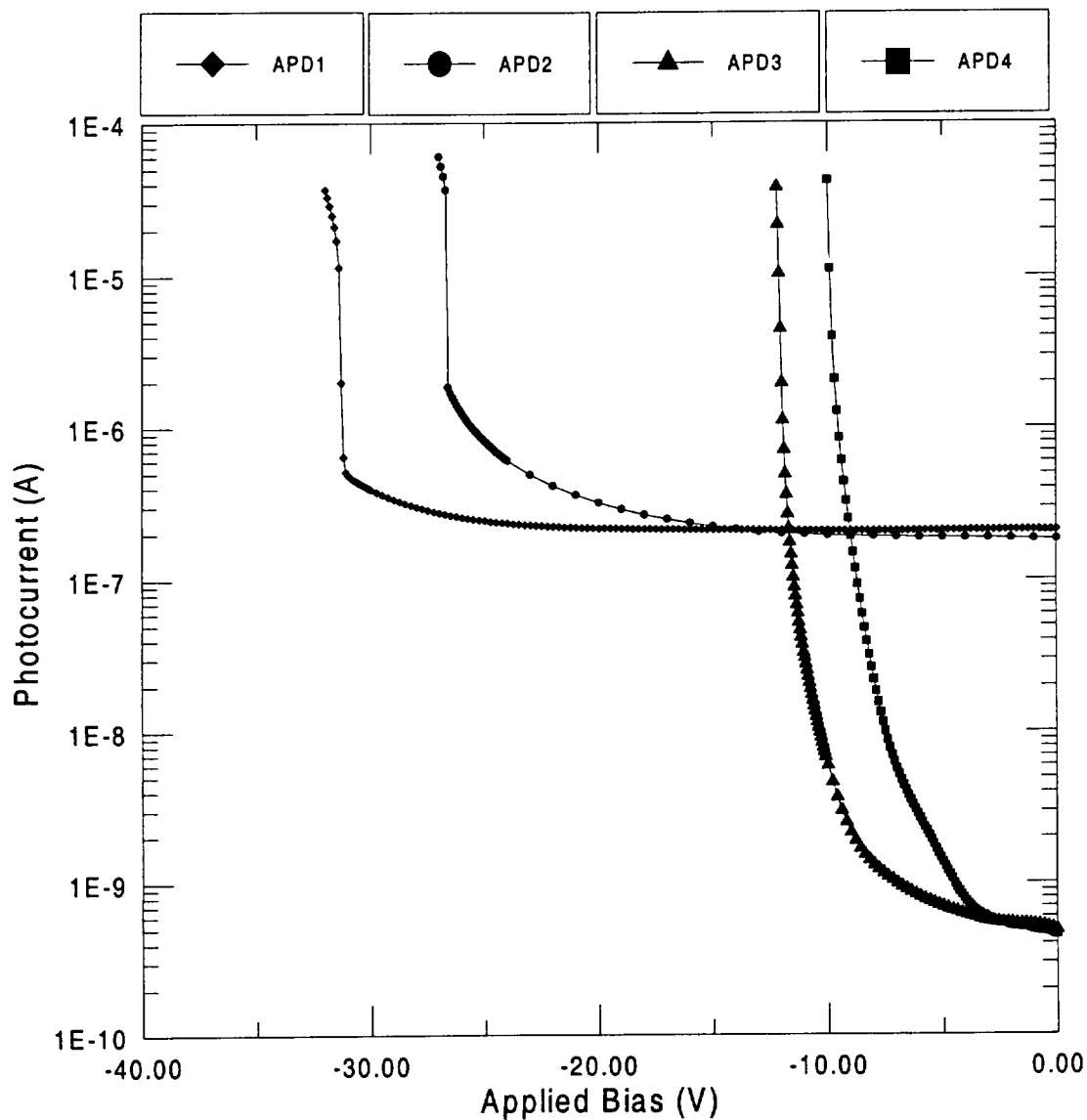


Figure 5-7: Experimental light IV curves obtained for all four 10-period doped-well MQW APDs for the same incident photon flux (electron injection)

CONCLUSIONS

Throughout this work, a detailed comparison of the gain and noise characteristics of a conventional and a doped well MQW APD was presented. The data obtained demonstrated a direct experimental evidence of structure induced preferential multiplication of electrons over holes. For the doped MQW APDs, the average gain per stage was calculated by comparing gain data with carrier profile measurements, and was found to vary from 1.03 at low bias to 1.09 near avalanche breakdown. This is in contrast to conventional PIN structures which show no gain in this regime. It was also shown that, as the bias was increased, the effect of the structure became less pronounced, and the MQW device was reduced to a conventional PIN structure. Similar studies of the bias dependence of the excess noise characteristics show that the low-voltage gain is primarily due to electron ionization in the MQW APDs, and to both electron and hole ionization in the PIN APDs. Our measurements of the doped MQW APD clearly showed that for low gains ($M < 6$), the ionization ratio is greatly enhanced ($k = \alpha/\beta = 10 - 50$) as compared to that in bulk GaAs ($k = 1.67$). At higher voltages, however, the value of k is reduced since the holes gain more energy from the applied electric field and are more likely to impact ionize.

It was also observed that surface recombination has a significant on the dark current behavior of an APD. The resulting leakage currents can have dramatic consequences on the sensitivities and attainable gain levels in a photodiode. As a result of surface treatment, dark currents at low bias were reduced to as low as 1 pA. The result of

this reduction in dark current was manifested in the structures' high gain performance which exceeded 10,000 in some APDs. By being able to reduce the dark currents, it was possible to maintain dark current levels well below those of the photocurrents. This made it possible to achieve and sustain high levels of gains well beyond the onset of junction breakdown.

The spectral response and quantum efficiencies for some of the structures were also calculated and modeled. The experimental data were very consistent with the theoretical models. The quantum efficiencies of the fabricated structures were relatively low because of the loss of photogenerated carriers due to recombination mechanisms in the diffusion layer. It was shown how it was theoretically possible to significantly increase the quantum efficiency of the devices through the introduction of a heavily doped p^{++} GaAs top layer. Such layers help to create a high-field region that will enhance the diffusion of photogenerated electrons toward the depletion region.

An investigation was also made into the impact of doping imbalances in doped-well MQW APDs on device IV and CV characteristics and how such an imbalance would affect the depletion properties of the APDs. Our theoretical models were in full agreement with the observed experimental data and have provided a good understanding of the physical processes that take place inside a doped MQW APD. These models have been used to interpret experimental IV and CV data and to determine the extent of depletion in APD devices. How these parameters were affected by the p- and n-doping imbalance in the structure was also determined. The model predicted that a doping mismatch as small as 10% could reduce the depletion layer by as much as 50%. It was

also shown how a large doping imbalance would cause the device to quickly reach avalanche breakdown in the depleted layers and would prevent full depletion.

The presence of undepleted regions was also shown to be one of the major causes of the slow time response in avalanche photodiodes. Partial depletion gave rise to a diffusion-limited transient response in doped MQW structures. This was demonstrated experimentally to be the case by examining the change in diffusion tail of the output pulse response of the devices as a function of applied bias. Fully depleted PIN structures showed a fast time response even at zero applied bias. The relationships between the depleted (undepleted) widths and the drift (diffusion) time response were used in conjunction with the sums of squares approximation to get an estimate of the time constants which limit the overall response of both the PIN and the MQW structures. The diffusion time constant for the MQW structure ($\tau_{\text{diff}} \sim 1.5$ ns) was found to be more than twice as large as that calculated for the PIN APD ($\tau_{\text{diff}} \sim 0.7$ ns) and confirms that the undepleted MQW structure was diffusion-limited.

PUBLICATIONS :

- [1] Menkara, H. M., Wagner, B. K., Summers, C. J., "Gain properties of doped GaAs/AlGaAs multiple quantum well avalanche photodiode structures", *Appl. Phys. Lett.*, Vol. 66, no. 14, pp. 1764-1766 April 1995.
- [2] Yun, I., Menkara, H. M., Wang, Y., Oguzman, I. H., Kolnik, J., Brennan, K.F., May, G. S., Summers, C. J., Wagner, B. K., "Reliability assessment of multiple quantum well avalanche photodiodes", *1995 IEEE International Reliability Physics Proceedings*. 33rd Annual, p. 200-204.
- [3] Menkara, H. M., Wagner, B. K., Summers, C. J., "Effect of variations in the doping profiles on the properties of doped multiple quantum well avalanche photodiodes", to be published in *Optical Materials*.
- [4] Yun, I., Menkara, H. M., Wang, Y., Oguzman, I. H., Kolnik, J., Brennan, K.F., May, G. S., Summers, C. J., Wagner, B. K., "The effect of manufacturing process sequence on the reliability of GaAs multiple quantum well avalanche photodiodes", to be published.
- [5] Menkara, H. M., B. K. Wagner, R. G. Benz II, Summers, C. J., "Computer Simulation and Design of Optoelectronic Devices", *Proceedings of the Second LAAS International Conference on Computer Simulation*, pg. 144, (1997).
- [6] Menkara, H. M., Bicknell-Tassius, R. N., Benz, R., Summers, C. J., "MBE Growth and Characterization of Doped Multiple Quantum Well Avalanche Photodiodes", *Journal of Crystal Growth*, 175/176, 983-989 (1997).

PRESENTATION :

Menkara, H. M., Wagner, B. K., Summers, C. J.

Effect of variations in the doping profiles on the properties of doped multiple quantum well avalanche photodiodes. First International Workshop on Materials for Optoelectronics, Sheffield Hallam University, Sheffield, U.K, August 1995.

Ph.D. Thesis:

H. Menkara, "A Comprehensive Analysis of the Physical Properties of Advanced GaAs/AlGaAs Junctions", Ga Tech's School of Physics, June 1996.

REFERENCES

- [1] K. Brennan, *IEEE J. Quantum Electron.*, QE-22, 1999 (1986)
- [2] F. Capasso, R. E. Nahory, M. A. Pollack, and T. P. Pearsall, *Phys. Rev. Lett.* 39, 723 (1977)
- [3] R. Chin, N. Holonyak, Jr., G. E. Stillman, J. Y. Tang, and K. Hess, *Electron. Lett.* 16, 467 (1980)
- [4] N. Yamamoto, K. Yokoyama, and M. Yamamoto, *Appl. Phys. Lett.* 62, 252 (1993).
- [5] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981), Chap. 2, p. 77.
- [6] R. U. Martinelli, T. J. Zamerowski, and P. Longway, *Appl. Phys. Lett.*, vol. 53, no. 11, p. 989, 1988.
- [7] K. R. Linga, G. H. Olsen, V. S. Ban, A. M. Joshi, and W. F. Kosonocky, *Journal of Lightwave Technology*, Vol. 10, no. 8, August 1992.
- [8] R. J. McIntyre, *IEEE Trans. Electron Devices* 13, 164 (1966).
- [9] M. Teich, K. Matsuo, and B. Saleh, *IEEE J. Quantum Electron.*, QE-22, No 8, 1184 (1986)
- [10] N. Hakim, B. Saleh, M. Teich, *IEEE Trans. Electron Devices*, vol. 37, No. 3, 599 (1990)
- [11] J. S. Marsland, *J. Appl. Phys.* 67, 1929 (1990)
- [12] J. S. Marsland, R. C. Woods, and C. A. Brownhill, *IEEE Trans. Electron Devices*, ED-39, 1129 (1992)

-
- [13] M. M. Hayat, B. E. Saleh, M. C. Teich, IEEE Trans. Electron Devices, ED-39, 546 (1992)
- [14] P. Aristin, A. Torabi, A.K. Garrison, H.M. Harris, and C.J. Summers, *Appl. Phys. Letters* 60, 85 (1992).
- [15] D.C. Raynolds, G. Leies, L. Antes, and R.E. Marburger, "Photovoltaic Effect in Cadmium Sulfide," *Phys. Rev.*, **96**, 533 (1954)
- [16] S.M. Sze, "Physics of Semiconductor Devices", Wiley, 1981.
- [17] P. Aristin, "Fabrication et Caracterization de Photodiodes a Avalanche a Puits Quantiques Multiples", Ph.D. thesis, January 1992.
- [18] S.R. Forrest, R.G. Smith, and O.K. Kim, IEEE J. Quantum Electron. QE-18 (1982) 2040.
- [19] R. Kuvas and C. A. Lee, J. Appl. Phys., 41, 1743 (1970)
- [20] J. N. Hollenhorst, IEEE J. Lightwave Technology, LT-8, 531 (1990)
- [21] B. C. Roy, and N. N. Chakrabarti, IEEE J. Lightwave Technology, LT-10, 169 (1992)
- [22] S.R. Forrest, and O.K. Kim, "Solid State Electronics", 26, 10, pp. 961-962 (1983)
- [23] M.S. Tyagi, "Solid State Electronics", 11, 99 (1968)
- [24] S.R. Forrest, and O.K. Kim, "Solid State Electronics", 26, 10, pp. 958-959 (1983)
- [25] M.J.O. Strutt, "Semiconductor Devices", Vol. 1, *Semiconductor and Semiconductor Diodes*, Academic, New York, 1966, Chap. 2.
- [26] Silvaco International, "Atlas II User's Manual", 1st Edition, July 1, 1993.
- [27] S. Selberherr, "Analysis and Simulation of Semiconductor Devices", Springer-Verlag, Wien-New York. 1984.
- [28] N. Yamamoto, K. Yokoyama, and M. Yamamoto, *Appl. Phys. Lett.* 62, 252 (1993).

PUBLICATIONS :

- [1] Menkara, H. M., Wagner, B. K., Summers, C. J., "Gain properties of doped GaAs/AlGaAs multiple quantum well avalanche photodiode structures", *Appl. Phys. Lett.*, Vol. 66, no. 14, pp. 1764-1766 April 1995.
- [2] Yun, I., Menkara, H. M., Wang, Y., Oguzman, I. H., Kolnik, J., Brennan, K.F., May, G. S., Summers, C. J., Wagner, B. K., "Reliability assessment of multiple quantum well avalanche photodiodes", *1995 IEEE International Reliability Physics Proceedings*. 33rd Annual, p. 200-204.
- [3] Menkara, H. M., Wagner, B. K., Summers, C. J., "Effect of variations in the doping profiles on the properties of doped multiple quantum well avalanche photodiodes", *Optical Materials*, Vol. 6, pp. 103-109, 1996.
- [4] Yun, I., Menkara, H. M., Wang, Y., Oguzman, I. H., Kolnik, J., Brennan, K.F., May, G. S., Summers, C. J., Wagner, B. K., "Effect of Doping on the Reliability of GaAs Multiple Quantum Well Avalanche Photodiodes", *IEEE Transactions on Electron Devices* Vol. 44, No. 4, p. 535-544, 1997.
- [5] Menkara, H. M., B. K. Wagner, R. G. Benz II, Summers, C. J., "Computer Simulation and Design of Optoelectronic Devices", *Proceedings of the Second LAAS International Conference on Computer Simulation*, pg. 144, (1997).
- [6] Menkara, H. M., Bicknell-Tassius, R. N., Benz, R., Summers, C. J., "MBE Growth and Characterization of Doped Multiple Quantum Well Avalanche Photodiodes", *Journal of Crystal Growth*, 175/176, 983-989 (1997).

PRESENTATION :

Menkara, H. M., Wagner, B. K., Summers, C. J.
Effect of variations in the doping profiles on the properties of doped multiple quantum well avalanche photodiodes. First International Workshop on Materials for Optoelectronics, Sheffield Hallam University, Sheffield, U.K, August 1995.

Ph.D. Thesis:

H. Menkara, "A Comprehensive Analysis of the Physical Properties of Advanced GaAs/AlGaAs Junctions", Ga Tech's School of Physics, June 1996.

SECTION III:

RELIABILITY MODELING AND PARAMETRIC YIELD PREDICTION OF GaAs MULTIPLE QUANTUM WELLS

Reliability Modeling and Parametric Yield Prediction of GaAs Multiple Quantum Well Avalanche Photodiodes

HDTV Project Report

Ilgu Yun and Gary S. May
School of Electrical and Computer Engineering
Georgia Institute of Technology

1. Overview

GaAs/AlGaAs multiple quantum well (MQW) avalanche photodiodes (APDs) are of interest as an ultra-low noise image capture mechanism for high definition systems. In this effort, the effect of various doping methods on the reliability of GaAs/AlGaAs MQW APD structures fabricated by molecular beam epitaxy (MBE) were investigated. Reliability was examined by accelerated life tests by monitoring dark current and breakdown voltage. Median device lifetime and the activation energy of the degradation mechanism are computed for undoped, doped-barrier, and doped-well APD structures. Lifetimes for each device structure were examined via a statistically designed experiment. Analysis of variance showed that dark current was affected primarily by device diameter, temperature and stressing time, and breakdown voltage depended on the diameter, stressing time and APD type. The undoped APD had the highest reliability, followed by the doped well and doped barrier devices, respectively.

To determine the source of the degradation mechanism for each device structure, failure analysis using the electron-beam induced current (EBIC) method was performed. EBIC analysis revealed some degree of device degradation caused by ionic impurities in the passivation layer, and energy-dispersive spectrometry (EDS) subsequently verified the presence of ionic sodium as the primary contaminant. However, sodium contamination alone did not account for the observed variation between the differently doped APDs. This effect was explained by dopant migration during stressing, which is verified by free carrier concentration measurements using the capacitance-voltage technique.

Since literally millions of APDs must be fabricated for HDTV imaging arrays, it was also critical to evaluate potential performance variations of individual devices in light of the realities of semiconductor manufacturing. This study also presented a systematic methodology for modeling the parametric performance of GaAs MQW APDs. The approach described requires a model of the probability distribution of each of the relevant process variables, as well as a second model to account for the correlation between this measured process data and device performance metrics. The availability of these models enables the computation of the joint probability density function required for predicting performance using the Jacobian transformation method. The resulting density function could then be numerically integrated to determine parametric yield. Neural networks were used as a tool for generating the models described above. In applying this methodology to MQW APDs, it was shown that using a small number of test devices with varying active diameters, barrier and well widths, and doping concentrations enables prediction of the expected performance variation of APD gain and noise in larger populations of devices. This approach compared favorably with the Monte Carlo technique, but consumed fewer computational resources.

2. Accomplishments

2.1. Reliability Modeling

In the image capture application, the image capture stage must have sufficient optical gain to enable very sensitive light detection, but at the same time, the gain derived during detection must not contribute additional noise. Various APD structures, including doped-barrier, doped-well, and undoped devices have been fabricated, and these structures were all considered as candidates for the high-definition system imaging application. The effect of the different doping techniques on device performance is critical. An

investigation of the relative advantages and disadvantages of each device structure as pertaining to long-term, low-noise performance was therefore warranted.

In the reliability study, accelerated life testing of undoped, doped barrier, and doped well APD device structures has been conducted with the objective of estimating long-term device reliability. Dark current and breakdown voltage were the parameters monitored. Degradation in these parameters was investigated via high temperature storage tests and accelerated life tests, and the results of these tests were used to estimate device lifetime by assuming an Arrhenius-type temperature dependence [1]. Using the median device lifetime and its standard deviation as parameters, a failure probability model of these devices was derived using a log-normal failure distribution [2].

Lifetimes for each device structure were examined via a statistically designed experiment. A comparison of the reliability of the various APD structures was then performed using the analysis of variance (ANOVA) technique [3]. Results of the ANOVA study that dark current was mainly dependent on device diameter, temperature and stress time. Breakdown voltage was primarily impacted by diameter, temperature and APD type. It was concluded that the undoped APD structure yielded devices that exhibited the highest reliability, followed by the doped well and doped barrier devices, respectively.

Following device stressing, an analysis was conducted to determine the failure mechanism. Potential failure mechanisms were evaluated using scanning electron microscopy (SEM) and EBIC method [4]. Based on SEM and EBIC analysis, the presence of ionic impurities contaminating the passivation layer at the junction perimeter was proposed as a potential failure mechanism. EDS [5] was subsequently used to identify ionic sodium as the source of contamination. However, all three device structures are passivated using the same procedure. Therefore, sodium contamination alone does not account for the observed variation between the differently APD device types. On the contrary, this result is explained by dopant migration during stressing, which is verified by the measurement of free carrier concentration before and after stressing using the capacitance-voltage (C-V) technique [6].

2.1.1. Accelerated Life Testing

The basic structure of the devices investigated (Figure 1) is that of a p-i-n diode where the intrinsic region is composed of the MQW superlattice structure. For the doped-barrier MQW APDs, the 1-3 μm thick GaAs/AlGaAs superlattice region consists of 25 periods of 200 \AA GaAs quantum wells separated by 800 \AA AlGaAs barrier layers. One complete period consists of a 300 \AA high-field AlGaAs region doped at $3 \times 10^{18} \text{ cm}^{-3}$, the 200 \AA undoped GaAs layer, and a 500 \AA undoped AlGaAs layer. In the doped-well devices, high electric fields are achieved in the narrow bandgap GaAs wells of the avalanche region by the introduction of 50 \AA thick p+ and n+ layers doped at $1.5 \times 10^{18} \text{ cm}^{-3}$ [16]. The undoped MQW APD design is similar, but with the MQW region replaced by a 2.5 μm intrinsic GaAs layer.

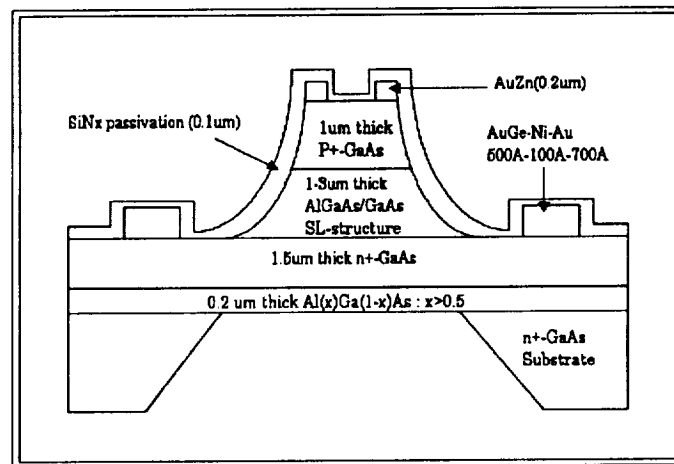


Figure 1. Cross section of AlGaAs/GaAs MQW APD.

Accelerated life tests for the three different APD structures were performed on several different devices of each type with a constant reverse current of 10 μA for 200 hours at three different ambient temperature levels: 100, 150 and 200 $^{\circ}\text{C}$. These conditions are summarized in Table 1.

Table 1. Accelerated Life Tests Conditions

Temperature [$^{\circ}\text{C}$]	Current [μA]	Number of Samples	Stress Time [hour]
100	10	4	200
150	10	6	200
200	10	6	200

To maintain a constant 10 μ A current, the reverse bias voltages for the doped-barrier, doped-well and undoped APD were approximately 8, 10 and 80 V, respectively. The activation energy for the failure mechanism and the average device lifetime were subsequently computed. It was assumed that the temperature dependence of the device failure rate (R) obeys the following Arrhenius law [1]:

$$R = R_o * \exp(-E_a / kT) \quad (1)$$

where R_o is a temperature-independent pre-exponential failure acceleration factor, E_a is the activation energy, T is the absolute temperature, and k is Boltzmann's constant. During these tests, dark current and breakdown voltage were measured at room temperature (300 °K) after high-temperature stressing. The breakdown voltage was obtained from the device I-V curve using the tangential line method. Typical breakdown voltages were 7.5 - 9 V for the doped-barrier APD, 10 - 12 V for the doped-well APD, and 70 - 85 V for the undoped APD. The devices were classified as failing when the dark currents at room temperature and 90% of the breakdown voltage exceeded 1 μ A.

Several observations were made as a result of the high temperature storage tests and accelerated life tests. First, unbiased baking of the APD samples resulted in significantly less degradation, which is demonstrated by a comparison of Figures 2 and 3. Dark current increases due to thermal overstress under bias for the doped-barrier devices were generally found to be exponentially dependent on the time of exposure to the reverse-bias field. The doped-well and undoped devices exhibited similar behavior. This fact is shown Figure 4(a), in which the dark current at a given reverse-bias voltage increases significantly as a function of stress time. On the other hand, breakdown voltage was shown to be nearly linearly dependent on stressing time, as shown in Figure 4(b).

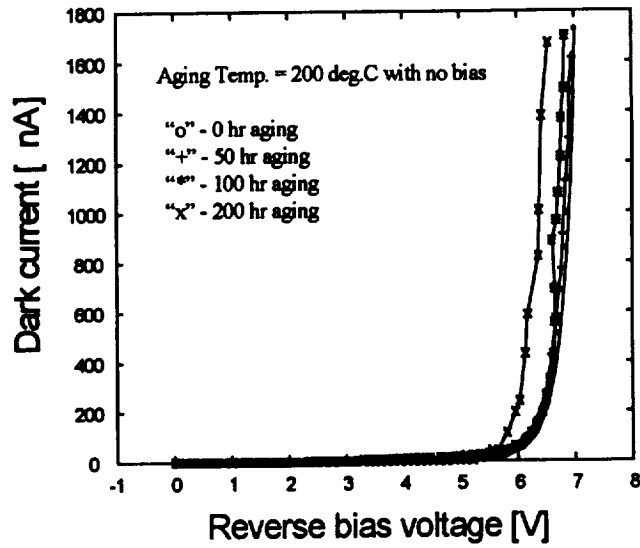


Figure 2. Room-temperature I - V curve of an doped-barrier APD sample after 0, 50, 100, and 200 hours of unbiased baking at 200 degrees C.

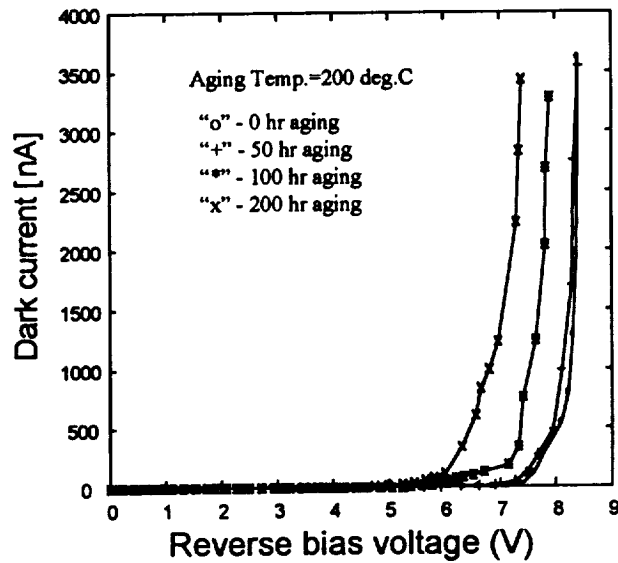


Figure 3. Room-temperature I - V curve of an doped-barrier APD sample after 0, 50, 100, and 200 hours of aging under bias at 200 degrees C.

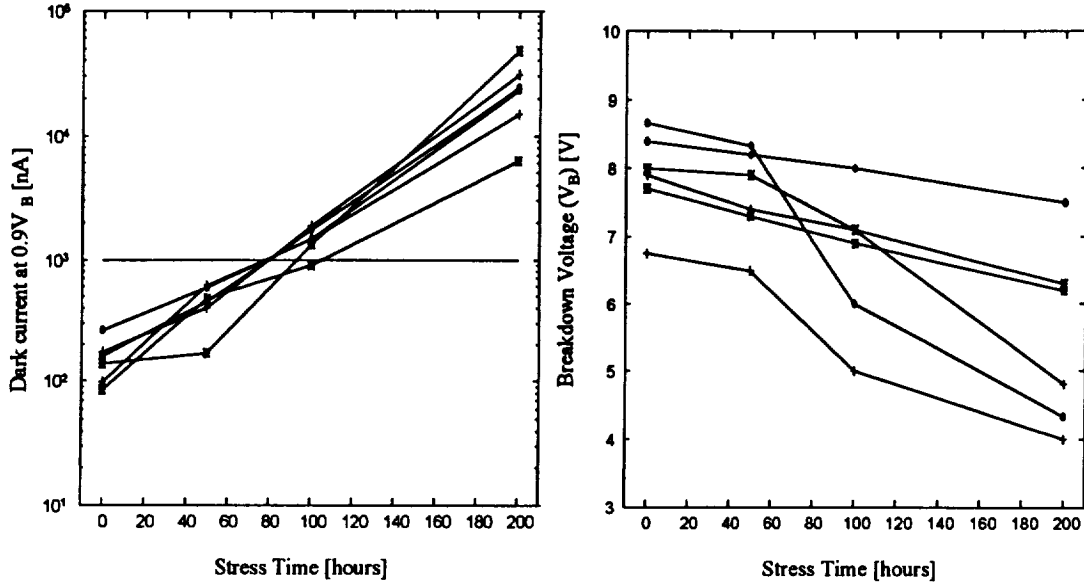


Figure 4. (a) Dark current and (b) breakdown voltage variations of doped-barrier APDs after accelerated life testing at 200 degrees C.

Figure 5 is an example of the percent of cumulative failures for the doped-barrier device versus the lognormal projection of the device time-to-failure after accelerated life testing. Although the sample size is small, the data appeared linear, which indicates that the failure mode is the wearout type. Failures obeyed the lognormal distribution relatively well. Median lifetimes for the doped-barrier devices at 100, 150, and 200 °C were estimated to be 1400, 250, and 78 hours, respectively, with a standard deviation of 1.84. For the doped-well APDs, median lifetimes at 100, 150, and 200 °C were estimated to be 4204, 315 and 86 hours, respectively, with a standard deviation of 1.94. Finally, in the undoped case, the median lifetimes at 100, 150, and 200 °C are estimated to be 8590, 495 and 84 hours, respectively, with a standard deviation of 2.13.

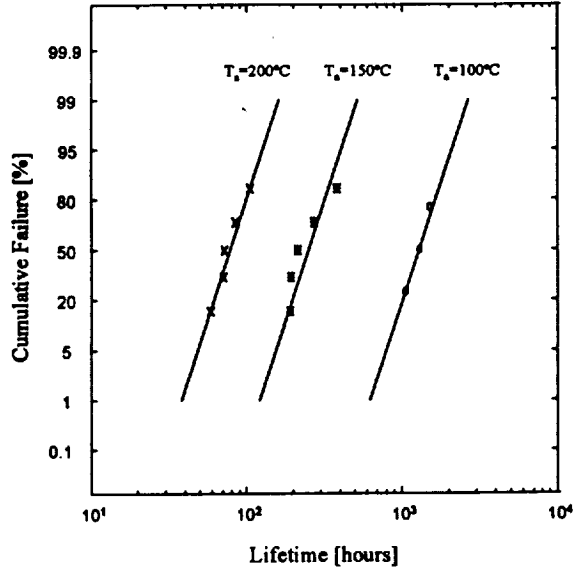


Figure 5. Lognormal projection of time-to-failure versus percent of cumulative failures for doped-barrier APDs after life testing at 100, 150, and 200 degrees C.

A sample Arrhenius plot of median lifetime as a function of reciprocal aging temperature is shown in Figure 6. From these plots, the thermal activation energy of the device aging process is computed to be 0.44, 0.60, and 0.71 eV for the doped-barrier, doped-well and undoped devices, respectively. Using these activation energy levels, the median APD lifetime for the doped-barrier device under practical use conditions can be estimated to be 3.7×10^4 hours (approximately 4.3 years) at room temperature, with a standard deviation of 116 hours. Lifetime estimates for the doped-well and undoped cases were 3.4×10^5 hours (approximately 39 years) with a standard deviation of 343 hours and 1.7×10^6 hours (approximately 197 years) at room temperature, with a standard deviation of 1031 hours. It is interesting to note that the doped-well APD, which is a complimentary structure of the doped-barrier APD, has a significantly longer median lifetime. A summary of life test results is shown in Table 2.

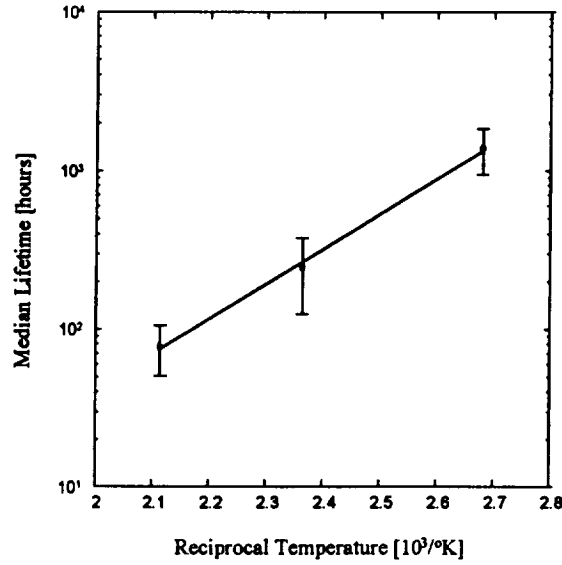


Figure 6. Arrhenius plot of median device lifetime for doped-barrier APDs as a function of reciprocal aging temperature.

Table 2. Summary of Life Test Results

Device Type	Activation Energy	Median Lifetime (at 300 °K)	Standard Deviation
Doped-barrier APD	0.44 eV	3.7×10^4 hours	116 hours
Doped-well APD	0.60 eV	3.4×10^5 hours	343 hours
Undoped APD	0.71 eV	1.7×10^6 hours	1031 hours

2.1.2. Performance Comparison of APD Structures

Statistical experimental design [3] was used to quantify the impact of each factor on APD reliability and to determine whether the differences between device structures were statistically significant. Due to the mixture of qualitative and quantitative input factors, a D-optimal experimental design with 24 runs was selected to identify the effect of input parameters on the measured responses [7]. The factors investigated in this experiment were device type, diameter of the active area, aging temperature, and stress time. A summary of these input factors is shown in Table 3. Dark current, breakdown voltage, and device lifetime were the measured responses.

Table 3 - Input Factors

Parameter	Values
APD Type	Doped-barrier APD (DB) Doped-well APD (DW) Undoped APD (UND)
Aging Temperature	100-200 °C
Stress Time	50-150 hour
Diameter	75-130 μm

A comparison of the various APD structures in terms of reliability was performed using the analysis of variance (ANOVA) technique. Experimental data was analyzed using the RS/Discover commercial software package [8]. Using this approach, it was verified that the different processes used to fabricate the three APD structures did indeed significantly impact the reliability of the devices. Using the ANOVA technique, the statistical significance of each input reflects the degree to which the parameter contributes to the variation of the measured responses. If the value of the statistical significance is less than 5%, then the input contribution to the variation of the measured response is considered significant with 95% confidence. Table IV shows the significance of each factor on the two responses.

Table 4 - Results of D-optimal Experiment

Factor	Statistical Significance		
	Dark Current (I_D)	Breakdown Voltage (V_b)	Device Lifetime (T_D)
Diameter	0.0132	0.0141	0.3151
Temperature	0.0009	0.2192	0.0008
Stress time	0.0013	0.0218	0.4128
APD type	0.2288	0.0001	0.0035

Results indicate that dark-current variation is affected primarily by diameter, temperature, stressing time, and to a lesser degree by the APD type. Breakdown-voltage variation depends on the diameter, stressing time and APD type. Interestingly, the stress temperature did not have a significant effect on the change in breakdown voltage. Device lifetime is impacted most significantly by stress temperature and APD type. From these results, it may be concluded that the doping process used in the fabrication of the APD structure has a profound impact on device reliability. Since the undoped devices exhibit the highest

degree of reliability, it can be assumed that doping, while enhancing device performance, makes the device less reliable.

2.1.3. Failure Analysis

Failure analysis on the thermally stressed devices was carried out using SEM and EBIC [4]. Prior to this analysis, the presence of contaminants in passivating nitrides at the junction was hypothesized as a possible cause for dark current increases during stressing. Defects causing device failure were detected in each type of device after life testing (Figure 7). (Similar results were observed in the doped well and undoped devices). Using EBIC analysis, local defects at the junction region change the electron-beam current indicating the reason for the device failure. Defects near the area of the junction were detected in the EBIC images, and nearly all the SEM images exhibit a similar pattern of defects in the exposed junction area as well. The only exception was the SEM image of an undoped device after life testing, which showed only a small defect in the junction.

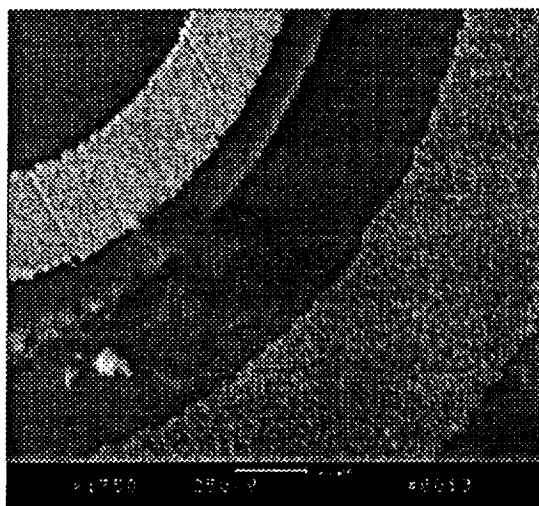


Figure 7. (a) SEM and (b) EBIC images of doped-barrier GaAs MQW APD after accelerated life testing at 200 degrees C.

From SEM and EBIC analysis of the degraded samples, it was determined that the dark current increase could be partially explained by the presence of ionic impurities or contamination in the silicon

nitride passivation layer at the junction perimeter. Such contamination generates a leakage path shorting the junction under an electric field. This hypothesis is supported by the fact that unbiased baking of the APD samples resulted in significantly less degradation, which is demonstrated by a comparison of Figures 2 and 3. It has been suggested that these type of defects occur at metal-rich precipitates, some of which occur at crystal dislocations [9-11]. The cause of the gradual reduction in breakdown voltage, on the other hand, is not known explicitly, but presumably involves the field-assisted and/or temperature-assisted drift of some impurity species or defects to localized sites in the pn junction.

A common contaminant for silicon nitride passivating films is ionic sodium. EDS was used to determine whether sodium was the source of contamination in these devices. In this case, EDS confirmed the presence of ionic sodium and verified that sodium is the primary contaminant (see Table 4). It is believed that this sodium originated from the APD processing environment or the personnel involved in fabrication. In addition, ionic potassium was detected in the doped-barrier device. (The significant amount of phosphorus detected in the undoped device was probably due to the etching of the mesa structure).

Table 4 - EDS Results for the Doped-barrier, Doped-well, and Undoped MQW APDs.

Element	Doped-barrier MQW APD	Doped-well MQW APD	Undoped MQW APD
	Weight [%]		
Na	13.68	18.39	14.48
In	-	3.72	1.21
P	4.38	-	21.79
Cl	10.41	-	-
K	11.31	-	-

Although ionic contamination is a plausible explanation for device degradation, this effect alone does not account for the statistically significant variations in lifetime among the differently doped APD structures. Since the same passivation process was applied to each structure, one would expect that each would have roughly the same lifetime if contamination were the sole cause of degradation. However, it was observed that the undoped devices were clearly more reliable, followed by the doped well and doped barrier

devices, respectively. Therefore, it was theorized that dopant migration might also play a significant role in the device degradation mechanism. This theory was investigated by analyzing dopant migration using C-V measurements to extract the free carrier concentration in the APD multiple quantum well region before and after life testing. C-V measurements were performed at 1 MHz using an HP4277A LCZ meter.

For the doped-barrier APD, the free carrier profile in the depletion region is shown in Figure 8. Before life testing, the depletion region width under a reverse bias near the breakdown voltage was approximately 0.195 μm . After life testing, the free carrier concentration significantly increased in the barrier region, and the depletion width decreased to 0.14 μm under reverse bias. Similarly, for the doped-well APD, the depletion region width under reverse bias was about 0.185 μm before life testing. After life testing, the free carrier concentration again increased, and the depletion width shrunk to 0.17 μm .

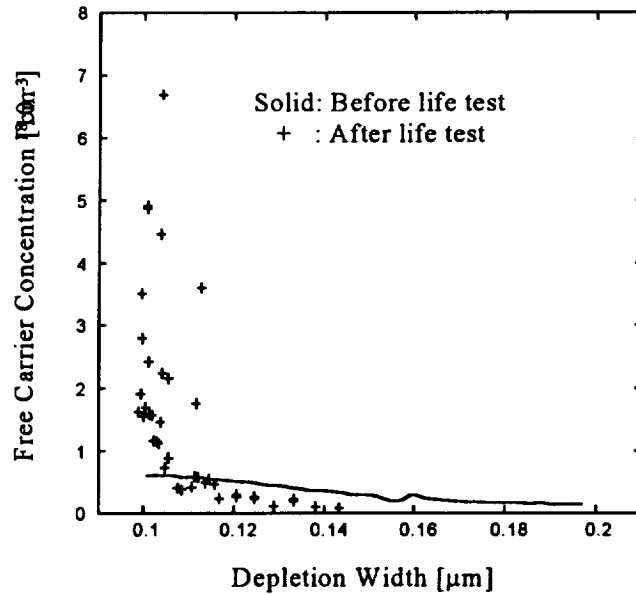


Figure 8. Free carrier concentration profile of doped-barrier APD before and after life testing.

The free carrier profile in Figures 8 is similar to those reported by Aristin et. al. for a doped-barrier MQW APD structure [12], who stated that as the doping concentrations in the barrier increase, dark current increases and breakdown voltage decreases. In the present investigation, the free carrier

concentrations increased in doped barrier layers after life testing as well, resulting in comparable increases in dark current.

From the results of the C-V measurements, it was hypothesized that during the life test, the thermally and electrically excited dopants obtain sufficient energy to migrate into the passivation layer, which caused an increase in free carrier concentration in this region. After entering the passivation layer, these dopants behaved similarly to positive surface charges. Because of the accumulation of positive charge, the depletion width was reduced and the electric field in the region where the p-n junction intersects the passivation layer was more intense. Dark current was increased by both the positive charge accumulation, as well as the intensified electric field in the narrow depletion region associated with the passivation layer. These increases accelerated the degradation of the device, eventually resulting in failure. The effect was more pronounced in the doped barrier devices since the observed shrinkage in the depletion region width is greater in these devices than in the doped-well APD.

2.2. Parametric Yield Modeling

Even in a defect-free manufacturing environment, random variations in the fabrication process will lead to varying levels of device performance. These manufacturing variations result from the fluctuation of various physical parameters (i.e., doping concentration, layer thickness, etc.), which in turn manifest themselves as variations in device performance metrics (such as gain or noise). This is due to the fact that these fluctuations influence the statistical distributions of device model parameters, which results in statistically varying performance characteristics in finished integrated circuits. Although small process fluctuations may not always cause catastrophic failures, they often prevent systems from meeting certain specifications. It is therefore crucial for circuit and device designers, as well as manufacturers, to account for statistical variations early in the design level, thereby aiding in production scheduling and planning.

This portion of the project developed a systematic methodology for modeling the parametric performance of GaAs MQW APDs. The approach described first requires a model which reflects the probability distribution of each of the relevant process variables. This model can be obtained directly from measured process data. A second model is then required to account for the correlation between this measured process data and device performance metrics. This can be derived either from the evaluation of analytical expressions relating process variables to performance or through device simulation. The availability of the above models enables the computation of the joint probability density function required for predicting performance using the Jacobian transformation method [13], which converts the process variable distributions to the device performance metric distributions. The resulting density function is then numerically integrated to determine parametric yield. Since they have demonstrated the capability of highly accurate function approximation and mapping of complex, nonlinear data sets, neural networks were used to generate these models [14].

2.2.1. Modeling and Simulation

APDs were only available in a very limited supply in this study (about ten each for the doped barrier, doped well, and undoped structure), so a thorough parametric study of gain and noise in these devices required other data in addition to that available directly from measurements. The objective of APD simulation was to use simulated data as a supplemental aid to experimental data for understanding the effect of variations in manufacturing parameters on APD gain and noise. Accurate simulation required that measured output data first be sufficiently calibrated with the simulation tool. Simulation of APD operation was performed using the ATLAS II device simulation package [15].

In the ATLAS II simulations, Newton's two-carrier method is used for solving Poisson's and the continuity equations. Impact ionization is modeled according to Selberherr [16]. Light I-V characteristics are modeled using a 1 mW/cm² monochromatic light source operating at 800 nm. From this simulation tool, dark current (I_D), photocurrent (I_L), and impact ionization rates for electrons and holes can be calculated. The multiplication gain is then given by:

$$M(V) = \frac{I_L(V) - I_D(V)}{I_{LO}} \quad (2)$$

where I_{LO} is the photocurrent at unity gain. The impact ionization rate ratio (k) is defined as the ratio of the electron to hole ionization rate (α_p/α_n).

To simplify the models and to reduce program execution time, the following assumptions were made regarding the simulated structures: 1) all devices have a rectangular geometric configuration; 2) only SRH and Auger recombination is considered (optical and surface recombination are ignored); 3) the p and n contacts are perfect ohmic contacts; 4) doping imbalances in the MQWs are constant throughout the entire structure; and 5) the effect of bandgap narrowing in AlGaAs is similar to that in GaAs. Figure 9 shows that the simulation results for a 10-period, doped-well MQW APD. The simulated I-V data matched the experimental data quite well, indicating that device characterization could justifiably be performed using data simulated by ATLAS II as a supplement to experimental data.

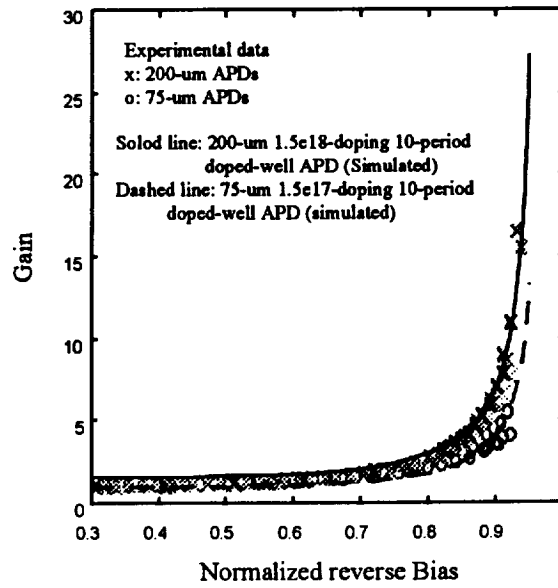


Figure 9. Measured and simulated results of gain vs. breakdown voltage for 10-period, doped-well MQW APDs.

The ATLAS II simulations were used to generate datasets to build neural network models which map the variations in device diameter, doping, and barrier width to device performance. Neural network modeling can be accomplished directly using the results obtained from the ATLAS II simulator. The use of neural networks to provide this mapping was motivated by the fact that neural net models can provide an accurate and efficient alternative to using ATLAS II directly. ATLAS II simulations typically took on the order of 30 minutes to run, whereas neural nets trained to mimic ATLAS provided results in fractions of a second.

Several simulations were performed using a systematic experiment designed to achieve sufficient coverage of the input parameter space, and the results of these simulations were used to train a neural network to model gain and noise index as a function of device diameter, barrier width, and the mean and standard deviation of the barrier (or well) doping. The gain

index was defined as the area under the plot of gain versus reverse bias up to the breakdown voltage. The noise index was defined by the electron-to-hole impact ionization rate ratio.

The four input factors varied in the gain and noise characterization simulations and their respective ranges of variation are shown in Table 5. These factors were selected due to their potential for variation in a manufacturing setting leading to possible impact on yield. The active diameter of an APD could vary due to photolithographic variations such as misalignment. The other three parameters (barrier width, mean doping, and doping standard deviation) are all subject to any fluctuations in the MBE system used to grow the APD superlattice. The ranges were selected to account for the variety of potential operating conditions used in device fabrication.

Table 5 - Input Factors for Gain and Noise Characterization

Parameter	Values
Active Diameter (D)	75-130 μm
Barrier Width (B)	200-800 \AA
The Mean doping (Dm)	10^{17} - 10^{18} cm^{-3}
The Standard deviation of doping(Ds)	10^{17} - 10^{18} cm^{-3}

The choice of an appropriate experimental design to capture variation in gain and noise over these ranges is important for systematically collecting data to be used for subsequent modeling. For this type of computer simulation study, Conover and Beckman have recommended the uniform design concept inherent in the Latin hypercube sampling (LHS) technique [16]. LHS is an extension of a stratified sampling procedure in which each input variable has all portions of its range represented in the design. The LHS design used in this study required 17 trials (see Table 6).

Table 6: Latin Hypercube Sampling Experimental Design

Run	x1 (D) [μm]	x2 (B) [\AA]	x3 (Dm) [10^{17}cm^{-3}]	x4 (Ds) [10^{17}cm^{-3}]
1	103	500	5.5	5.5
2	101	360	7.0	6.4
3	123	520	6.4	4.6
4	115	280	4.6	4.0
5	104	720	2.8	5.2
6	119	590	7.6	8.2
7	97	320	8.8	3.4
8	93	480	1.0	7.0
9	86	600	4.0	1.6
10	112	640	1.6	8.8
11	82	240	5.2	2.8
12	79	680	2.2	5.8
13	126	440	3.4	9.4
14	90	760	5.8	10.0
15	130	400	9.4	7.6
16	108	200	8.2	1.0
17	75	800	10.0	2.2

Neural networks possess the capability of learning complex relationships between groups of related parameters [18]. Such learning capabilities are attributed to the fact that neural networks, possessing many simple parallel processing units (called "neurons"), crudely resemble the architecture of the human brain. The most popular method of training feed-forward neural networks is the error back-propagation (BP) algorithm. BP networks consist of several layers of neurons which receive, process, and transmit critical information regarding the relationships between the input parameters and corresponding responses (Figure 10). Each neuron contains the weighted sum of its inputs filtered by a nonlinear sigmoidal transfer function. These networks incorporate "hidden" layers of neurons which do not interact with the outside world, but assist in performing classification and feature extraction tasks on information provided by the input and output layers. Two BP neural nets have been trained to

predict APD gain and noise. Inputs to the gain and noise neural network models are the parameters listed in Table 5.

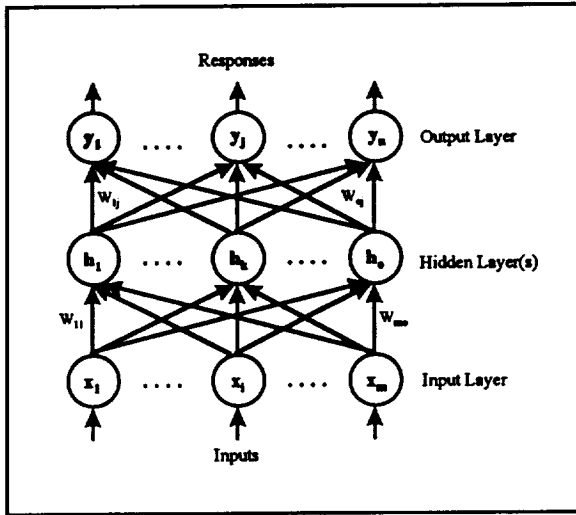


Figure 10. Typical feed-forward neural network.

2.2.2. Parametric Yield Prediction

Usually, it is assumed that these manufacturing parameters in Table 5 will vary according to the normal distribution. However, this may not always be the case [19]. Several commonly occurring distributions in semiconductor device fabrication are shown in Figure 11. These deviations from the ideal Gaussian shape can sometimes appear in IC fabrication. Each of these were considered as possible distributions in APD fabrication in order to accurately characterize fluctuations in parametric yield.

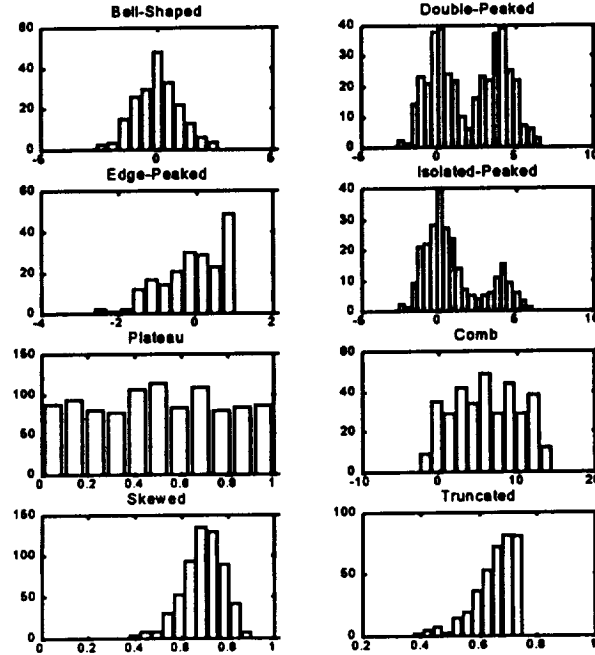


Figure 11. Possible distributions occurring in semiconductor device fabrication.

The histograms described above provide models of marginal probability density functions (pdfs) for each device parameter. These marginal pdfs are related to the joint probability density function for all parameters as follows [17]:

$$f_1(x_1) = \int \dots \int f(x_1, x_2, \dots, x_n) dx_n \dots dx_2 \approx \sum_{x_2} \dots \sum_{x_n} f(x_1, x_2, \dots, x_n) \quad (3)$$

where $f_1(x_1)$ is the marginal pdf for parameter x_1 and $f(x_1, x_2, \dots, x_n)$ is the joint pdf for n different device parameters. In this work, the x_i 's are the manufacturing parameters. Multiple integrals of the joint pdf using Eqn. (3) provide probability information along several dimensions in the same way that integrating a marginal pdf gives the probability of finding a single variable in a given interval.

The joint pdf is found by determining the relative frequency of device performance along several dimensions. This is accomplished by partitioning the device parameter space into divisions with appropriate granularity, counting the number of devices in each category, and dividing by total number of devices measured. Then, the last step ensures that joint pdf is normalized. This procedure can be extended to as many dimensions as desired, and the resulting hypersurface likewise approximates the multidimensional joint pdf.

Since the exact form of the manufacturing parameter distributions is difficult to predict, the assumption of normal behavior may be incorrect. To circumvent this difficulty, neural networks were used as a mechanism to encode the functional form of the overall joint parameter distribution directly from measured (or simulated) data.

Once the joint pdf of the device parameters was computed, the next step was to derive the joint pdf for functions of these parameters. For example, if the joint pdf of active diameter (A) and barrier width (B) is known, we would like to use this information to calculate the joint pdf of device performance characteristics such as gain or noise index, since each of these performance measures is a function of A and B. Consider two sets of random variables X_j (representing the manufacturing parameters) and Y_i (representing the performance metrics), where the Y_i 's are functions of the X_j 's:

$$x_1 = A; x_2 = B; y_1 = G; y_2 = N \quad (4)$$

The functional relationship between the manufacturing process variables and performance metrics can be expressed as:

$$\begin{aligned} y_1 &= H_1(x_1, x_2) \\ y_2 &= H_2(x_1, x_2) \end{aligned} \quad (5)$$

where H_1 and H_2 are continuous, differentiable functions. Now x_1 and x_2 can be solved in terms of y_1 and y_2 to obtain:

$$\begin{aligned} x_1 &= G_1(y_1, y_2) \\ x_2 &= G_2(y_1, y_2) \end{aligned} \quad (6)$$

where G_1 and G_2 are also continuous and differentiable. The joint pdf of random variables Y_1 and Y_2 , $u(y_1, y_2)$, is given by:

$$u(y_1, y_2) = f(x_1, x_2) |J(y_1, y_2)| \quad (7)$$

where $f(x_1, x_2)$ is the joint pdf of x_1 and x_2 , and $J(y_1, y_2)$ is the Jacobian transformation. The Jacobian is given by the following determinant:

$$J(y_1, y_2) = \begin{vmatrix} \frac{dx_1}{dy_1} & \frac{dx_1}{dy_2} \\ \frac{dx_2}{dy_1} & \frac{dx_2}{dy_2} \end{vmatrix} \quad (8)$$

Recall that the joint pdf of the manufacturing parameters, $f(x_1, x_2)$, is available from the previously obtained neural network models of the joint parameter density.

Once $u(y_1, y_2)$ has been calculated from Eqn. (7), then the marginal densities of the device performance metrics (gain index or noise index) is calculated as follows:

$$\begin{aligned} I_1(y_1) &= \int u(y_1, y_2) dy_2 \approx \sum_{y_2} u(y_1, y_2) \\ I_2(y_2) &= \int u(y_1, y_2) dy_1 \approx \sum_{y_1} u(y_1, y_2) \end{aligned} \quad (9)$$

where $I_1(y_1)$ and $I_2(y_2)$ are the marginal pdfs of the performance characteristics and the numerical integration is performed using the trapezoid rule. The parametric yield of the circuit with respect to a given performance measure is then derived from the marginal pdfs as:

$$PY_i = \int_a^b I_i(y_i) dy_i \approx \sum_{k=0}^{n-1} \left\{ I_i \left(a + \frac{k(b-a)}{n} \right) + I_i \left(a + \frac{(k+1)(b-a)}{n} \right) \right\} \left[\frac{b-a}{2n} \right] \quad (10)$$

where a and b represent the limits of integration surrounding regions of interest and PY_i provides the probability of the device satisfying a particular performance criterion. To evaluate this integral numerically, the interval $[a,b]$ is divided into n distinct segments. Using this methodology, the parametric yield of gain or noise can be predicted based on the variation of the manufacturing parameters.

2.2.3. Results and Discussion

The neural network models for gain and noise index were established from 17 ATLAS II simulations from the LHS design. Two three-layer neural networks with four inputs, five hidden neurons, and a single output were used. The networks were trained using ObOrNNS ("Object-Oriented Neural Network Simulator"). Prediction errors for each model were evaluated using a validation set consisting of four randomly selected runs from the LHS design. The training errors were 1.3% and 1.0%, respectively, and the prediction errors were 1.8% and 3.2%, respectively.

Based on the results of the neural network modeling, the effect of the various manufacturing parameters on gain and noise index can be quantitatively investigated. Figure 12 shows 3-D contour plots of gain and noise index versus active diameter and the mean value of the doping concentration. In each case, barrier width and the standard deviation of doping concentration remain constant at their mid-range value. It is evident that increasing the mean doping concentration results in higher gain. In addition, increasing the active diameter of the APD along with the mean doping concentration results in a higher noise index. These results occur partly due to the fact that increasing the doping concentration can cause more carrier multiplication during the avalanche process, which can increase impact ionization rate ratio

(k). As k increases, both the gain and noise index increase as well. These results are in agreement with experimental measurements performed by Aristin, et al. [12].

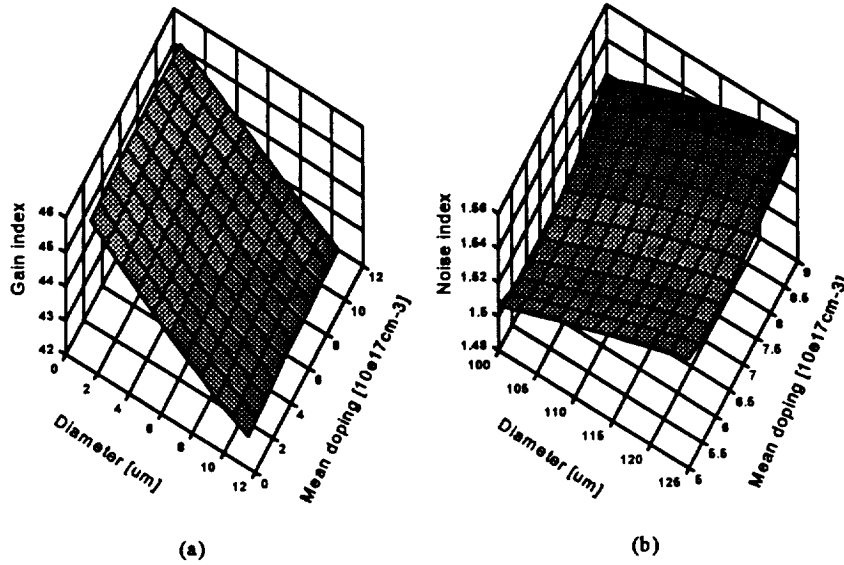


Figure 12. Contour plots of neural network models of: (a) gain index and (b) noise index as a function of mean doping concentration and device diameter. Barrier width and doping standard deviation are set to their mid-range values.

In addition to the above models, "inverse" neural network models are also needed for calculating the parametric yield using the procedure described above. These inverse models have been constructed by simply training a network with four inputs (gain index, noise index, and two "dummy" variables) and four outputs (the process parameters in Table 5). The two "dummy" variables are not directly involved in the parametric yield calculations, but are necessary to derive a proper Jacobian determinant [20].

To construct a joint density function for the four processing parameters, four different statistical distributions from those shown in Figure 11 were selected, and random numbers were generated according to these four distributions using MATLAB. The arbitrarily selected distributions were the bell-shaped, truncated, plateau, and combed distribution for device diameter, barrier width, mean value of doping concentration, and standard deviation of doping

concentration, respectively. Using data derived from these distributions as training data, a back-propagation neural network with a 4-9-1 (input-hidden-output) architecture was used to model the joint density function for all four input variables. To verify this model, the marginal density functions for each input variable were reproduced as shown in Figure 13. As this figure shows, the marginal distribution of each input parameter was well-matched with the neural network predictions.

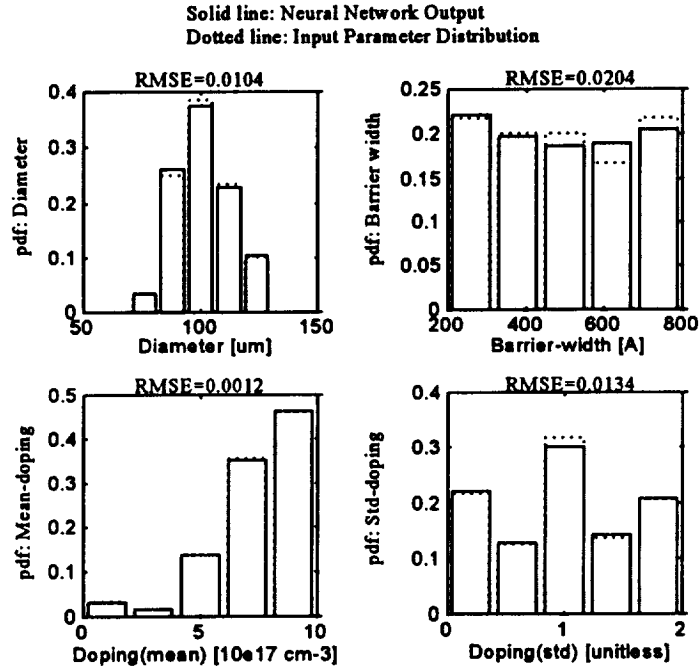


Figure 13. Predictions of joint pdf neural network model.

To calculate the parametric yield using the joint density function, the Jacobian determinant must be calculated. The derivatives required for the Jacobian matrix were estimated using a difference equation which computes the change in the input quantity with respect to a 5% deviation in the output parameter of interest. Following the computation of the Jacobian determinant, parametric yield may be calculated using Eqn. (7). Figure 14 shows the resulting distribution of gain and noise index.

Figure 14. Final distributions of gain and noise index. A 1- σ confidence interval has been indicated on the Monte Carlo result using considering different distributions (which are the target values).

To validate this approach, these results can be compared with the Monte Carlo method. Toward that end, simulations consisting of 20,000 randomly generated instances of data were used to calculate parametric yield using the Monte Carlo technique. These randomly generated data sets were fed into the neural network models for gain and noise to calculate the device response. Using the output of the neural network model for each instance of input data, the distribution of gain and noise can be calculated. Two different Monte Carlo simulations were performed. The first approach assumed that all input parameters were independent and normally distributed, ignoring any correlations which might exist between the input parameters. The second Monte Carlo simulation, however, made use of the different input distributions provided in Figure 13. The Monte Carlo results are also shown in Figure 14.

As expected, Figure 14 shows that the Monte Carlo method performed without considering the variety of input parameter distributions cannot predict parametric yield accurately. For example, if the multinormal Monte Carlo approach is used, the number of devices achieving a gain index between 45–47 is severely overestimated. Likewise, the number of devices with a gain index from 47–49 is underestimated. In either case, this approach gives misleading information about the effect of the APD manufacturing parameters on device performance.

On the other hand, the newly proposed methodology for parametric yield calculation is comparable to results achieved using the Monte Carlo method that does consider different

(and potentially correlated) input distributions, but with significantly fewer simulations. Although some computational overhead is incurred in deriving the neural network pdf and jpdf models (on the order of a few hundred simulations for generating the forward and reverse neural network models and for performing the Jacobian transformation numerically, depending on the granularity of the jpdf model), these models only need to be derived once. In contrast, the Monte Carlo procedure will always require a large number of simulations. Furthermore, it is also asserted here that since the Monte Carlo method uses the same distribution for each input parameter and does not account for possible correlations between parameters, some degree of accuracy is inherently lost. By modeling the input pdfs and their joint pdf directly, the proposed method overcomes this shortcoming.

3. Impact

This portion of the HDTV project has first provided accelerated life tests of doped-barrier, doped-well, and undoped AlGaAs/GaAs multiple quantum well avalanche photodiodes from the viewpoint of evaluating long-term reliability. From the life test results, the activation energy of the degradation mechanism and median lifetime of these devices was determined. Using the ANOVA technique, a comparison of the reliability of the various APD structures was then undertaken. Based on this investigation, it was concluded that the doping process used in the multiple quantum well APD fabrication has a significant effect on device reliability. The undoped APD structure yielded devices that exhibited the highest reliability, followed by the doped well and doped barrier devices, respectively. Subsequent failure analysis using the SEM and EBIC methods clarified that the dark current increase was in part brought about by the presence of ionic contaminants in the passivation layer at the junction perimeter that generate a leakage path which shorts the junction under the effect of electric field. EDS

analysis identified the primary contaminant as ionic sodium. In addition, dopant migration under stress was theorized as a means to explain the observed reliability differences between the device structures. This dopant migration was investigated using C-V measurements, which verified that the redistribution of free carriers after stress is indeed a plausible explanation for reliability differences.

This effort has also provided a systematic methodology which uses the unique capabilities of neural networks to model the parametric performance of GaAs MQW APDs. It was shown that using a small number of test devices enables accurate prediction of the expected performance variation of APD gain and noise in larger populations of devices. The methodology presented could also be utilized in other high-volume manufacturing applications. This approach could thus provide device designers with the ability to understand the manufacturability of various design options and enables process engineers to determine the consequences of process modifications. This will potentially allow parametric yield estimation prior to high-volume production in order to evaluate the impact of design decisions and process capability.

4. Students

Dr. Ilgu Yun graduated in December, 1997. Dr. Yun is currently with the Electronics and telecommunications Research Institute in Korea.

5. Future Work

Neural network modeling for parametric yield of GaAs MQW APDs has been accomplished in this project. However, it was a nontrivial effort to obtain the inverse neural network models required for estimating the parametric yield based on device lifetime. Usually one hidden layer is enough to build a neural network model, but two

hidden layers were required to build accurate models for device lifetime modeling. One task that needs to be performed in the future is neural network model optimization. Genetic algorithms have been successfully used for parameter optimization of neural networks and recipe synthesis in semiconductor manufacturing process [21]. Thus, if genetic algorithms are applied for neural network optimization in this application, more accurate and efficient neural network models should be achieved.

Another task to be pursued in the future is to utilize the methodology for calculating parametric yield described in this thesis to semiconductor devices fabricated by other certain manufacturing processes. For instance, parameter variation for the high electron mobility transistor (HEMT) fabricated by MBE can be investigated using this methodology.

In addition, this methodology can also be applied for modeling circuit performance. For example, the parameter variation for CMOS circuit can be predicted using this approach. The circuit variables can be modeled by measured data or SPICE circuit simulations. It was shown that this approach is superior to the prevailing Monte Carlo method by reducing the computation load and relaxing the assumption of a specific statistical distribution. If parametric yield prediction for newly developed circuits using a given manufacturing process is needed, this new methodology can predict parametric yield with a small number of test structures prior to high-volume manufacturing in order to evaluate the impact on manufacturability of both design specifications and process capability.

6. Publications

- [1] I. Yun, H. Menkara, K. Brennan, G. May, C. Summers, and B. Wagner, "Reliability Assessment of Multiple Quantum Well Avalanche Photodiodes," *Proc. 1995 International Reliability Physics Symposium*, April, 1995, pp. 200-204.
- [2] I. Yun, H. Menkara, K. Brennan, G. May, C. Summers, and B. Wagner, "Effect of Doping on the Reliability of GaAs Multiple Quantum Well Avalanche Photodiodes" *IEEE Trans. Elec. Dev.*, vol. 44, no. 4, pp. 535-544, April, 1997.
- [3] I. Yun and G. May, "Evaluating the Manufacturability of GaAs/AlGaAs Multiple Quantum Well Avalanche Photodiodes Using Neural Networks," *Internat'l Elec. Manufac. Tech. Symp.*, Oct., 1997, pp. 105-112.
- [4] I. Yun and G. May, "Parametric Yield Estimation of GaAs/AlGaAs Multiple Quantum Well Avalanche Photodiodes Using Neural Networks," *Intelligent Engineering Systems Through Artificial Neural Networks*, vol. 7, pp. 845-853, (C.H. Dagli, Ed.), New York: ASME Press, 1997.
- [5] I. Yun and G. May, "Parametric Manufacturing Yield Modeling of GaAs/AlGaAs Multiple Quantum Well Avalanche Photodiodes," *IEEE Trans. Semi. Manufac.*, vol. 12, no. 2, pp. 238-251, May, 1999.

REFERENCES

- [1] W. Joyce, K. Liou, F. Nash, P. Bossard, and R. Hartman, "Methodology of Accelerated Aging," *AT&T Tech. Journal*, vol. 64, no. 3, March, 1985.
- [2] F. Nash, *Estimating Device Reliability*, Boston: Kluwer, 1993.
- [3] G. E. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*, New York: Wiley, 1978.
- [4] E. Pollino, *Microelectronic Reliability*, vol. 2, Boston: Artech House, 1989.
- [5] T. B. Vander Wood, "Identification of Particulate Contaminants in IC Manufacture," *Solid State Tech.*, vol. 28, no. 8, Aug., 1985.
- [6] P. Blood and J. W. Orton, *The Electrical Characterization of Semiconductors: Majority Carriers and Electron States*, London: Academic Press, 1992.

- [7] Z. Galil and J. Kiefer, "Time- and Space-Saving Computer Methods, Related to Mitchell's DETMAX, for Finding D-Optimum Designs," *Technometrics*, vol. 22, Aug., 1980.
- [8] *RS/Discover User's Guide*, BBN Software Products Corporation, June 1988.
- [9] H. Sudo, M. Suzuki, and N. Miyahara, "Observation of Surface Degradation Mode of InP/InGaAs APDs During Bias-Temperature Test," *IEEE Elec. Dev. Lett.*, vol. EDL-8, no. 9, Sept., 1987.
- [10] H. Sudo and M. Suzuki, "Surface Degradation Mechanism of InP/InGaAs APDs," *IEEE J. Lightwave Tech.*, vol. 6, no. 10, Oct. 1988.
- [11] Y. Kuhara, H. Terauchi, and H. Nishizawa, "Reliability of InGaAs/InP Long-Wavelength p-i-n Photodiodes Passivated with Polyimide Thin Film," *IEEE J. Lightwave Tech.*, vol. LT-4, no. 7, July, 1986.
- [12] P. Aristin, A. Torabi, A. Garrison, H. Harris, and C. Summers, "New Doped Multiple Quantum Well Avalanche Photodiode: The Doped Barrier Al(0.35)Ga(0.65)As/GaAs Multiple Quantum Well Avalanche Photodiode," *Appl. Phys. Lett.*, vol. 60, no. 1, pp. 85-87, Jan., 1992.
- [13] D. Gibson, R. Poddar, G. May, and M. Brooke, "Statistically Based Parametric Yield Prediction for Integrated Circuits," *IEEE Trans Semi. Manufac.*, vol. 10, no. 4, pp.445-458. Nov. , 1997.
- [14] D. Gibson, R. Poddar, G. May, and M. Brooke, "Using Neural Networks to Estimate Probability Density Functions," *Proc. 1995 World Congresss on Neural Networks*, vol. I, pp. 122-127, July, 1995.
- [15] *ATLAS II User's Manual*, Silvaco International, 1993.
- [16] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, New York: Springer-Verlag, 1984.
- [17] M. McKay, W. Conover, and R. Beckman, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Outputs from a computer code," *Technometrics*, vol. 21, pp. 239-245, 1979.
- [18] G. May, "Manufacturing ICs the Neural Way," *IEEE Spectrum*, vol. 31, no. 9, pp. 47-51, Oct., 1994.
- [19] M. Page, "Analysis of Non-normal Process Distributions," *Semiconductor International*, vol. 17, no. 12, pp. 88-96, Oct., 1994.

- [20] W. Hines and D. Montgomery, *Probability and Statistics in Engineering and Management Science*, New York: Wiley, 1980.
- [21] S. Han and G. May, "Using Neural Networks to Perform PECVD Silicon Dioxide Recipe Synthesis via Genetic Algorithms," *IEEE Trans. Semi. Manufac.*, vol. 10, no. 2, pp. 279-287, May, 1997.

SECTION IV:

THEORY AND MODELING OF
ACT IMAGER CHIP

Final Report – NASA HDTV Project

Principal Investigator: Prof. Kevin F. Brennan

Affiliation: School of ECE, Georgia Tech

I. Accomplishments

Prof. Brennan coordinated the theoretical and modeling effort for the project. The main mission of the theory and modeling group was to assist in the design, optimization and evaluation of each component of the imager chip. It is first useful to outline the design of the ACT imager chip. The chip can be divided into three different components. These are 1) the image capture component, 2) charge accumulation and storage component, and 3) the ACT readout mechanism. During the course of this project, the theory and modeling team made important seminal contributions to each of these three parts of the design. In this section, we review the theory and modeling work performed on each of these components of the HDTV imager chip.

A. Image Capture Component

We examined two general image capture devices. The first and most important device considered was an avalanche photodiode, APD. An APD is particularly attractive to HDTV imaging since it provides internal gain necessary in overcoming the noise of the following electronics in the camera. As is well known, an APD provides gain through carrier multiplication via impact ionization events. Accurate modeling of an APD thus requires accurate modeling of impact ionization. Much of the early work on impact ionization was based on simplified analytical models of the process. Though these techniques were found to be useful in some instances, they have limited validity. More advanced numerical models of impact ionization are necessary to render an accurate

accounting of the mechanism. For this reason, during this project, we developed *new, numerical techniques for determining the impact ionization transition rate*.

The first approach adopted in our work to improve the formulation of the impact ionization transition rates was to perform a numerical calculation of the rate. We calculated the interband impact ionization transition rate for both silicon [1] and GaAs [2]. The full details of the analysis are discussed in references 1 and 2 and are included at the end of this report. The salient features of the technique and our results will nevertheless be discussed here for completeness. The transition rate was determined using Fermi's golden rule from a two-body screened Coulomb interaction assuming energy and momentum conservation. The transition rate was calculated by numerically integrating the over the full Brillouin zone including the full details of the band structures. It was found that the transition rate depends strongly on the initiating electron wavevector and that the transition rate is greatest in both silicon and GaAs for electrons within the second conduction band as compared to the first conduction band. Though this approach is highly accurate, it is numerically intensive requiring significant computational resources. In order to find a more computationally efficient algorithm we examined an alternative approach by which the computational demands of the fully numerical formulation can be mitigated yet high accuracy can be retained.

A modified analytical relationship for the impact ionization transition rate that includes the effects of the band structure for the initial state known as the Quade formula [3] was then examined. We succeeded in rederiving this formulation and in comparing its predictions to those made using either a more simplified model, the Keldysh formula, or numerical techniques. The results of our investigation are reported in reference [4] which

is included at the end of this report. It was found that the Quade model overcomes some of the limitations of the Keldysh formula and the numerical methods but at the expense of new parameterization. It was determined that the Quade model can qualitatively reproduce results similar to those obtained with the numerical model yet with far greater computational efficiency. Nevertheless, there exist some important limitations to the Quade formulation. These are that it does not accurately reproduce the quantum yield data for bulk silicon, it requires determination of a new parameter, related physically to the overlap integrals of the Bloch states which can only be adjusted by comparison to experiment, and fails to account for any wavevector dependence of the overlap integrals. As such we determined that use of the Quade formula is not justified in most situations. Instead, it is necessary to proceed with numerically generated values for the ionization transition rates.

On the basis of the above calculations, we were then able to refine our models for impact ionization and incorporate these into device simulators to study APD device performance. The simulators were then used to design and optimize different APD designs. *Using these simulation tools, we designed a new APD structure, the delta-doped p-I-n quantum well APD.* The full details of the device and its predicted workings are described in the paper of reference [5]. This new structure and its earlier derivatives [6] formed the basis for the light capture component for the ACT imager chip. Thus the theory and modeling effort produced a new APD design and assessed its working capabilities.

In addition to our work on APD structures, we also examined the workings of metal-semiconductor-metal, MSM, photodetectors as possible candidates for the

photodetector stage of the device. The results of our investigation are discussed in detail in references [7-9] which are enclosed at the end of this report. Our work on this portion of the project centered on developing a detailed model and utilizing it to determine the workings of an MSM device. We explored different designs utilizing GaAs/AlGaAs and GaInAs/AlInAs. Various structures were examined and compared in terms of their frequency response and responsivity.

B. Charge Accumulation and Storage Component

The second component within the ACT imager chip is the charge accumulation and storage device. In the present design we considered storing the accumulated charge within an open circuited APD device. The basic operation is as follows. The APD is configured in the charge storage mode where the device can collect charge over the entire integration period rather than just during the short sample/read-out time. In this mode the diode is initially biased to a potential somewhat below the breakdown voltage yet high enough that impact ionization is prevalent. Following the reset period the external bias is removed, the diode is physically isolated from the circuit and, over time, recovers back to equilibrium. The diode recovers to equilibrium through a combination of effects such as generation, capture of signal and background radiation, and leakage to the external read-out circuit. During the first stage of recovery, the charge added to the diode can be multiplied within the high-field depletion region of the device and is stored within the internal capacitance of the junction. We have examined the operation of an APD in the charge storage mode using our advanced numerical simulator. The results of that investigation are reported in reference [10] which is enclosed at the end of this report.

We investigated the behavior of silicon based APDs operated in the charge storage mode. The different issues investigated included the device sensitivity to the input photo-current including the self-quenching of the diode and its limitations in sensing low light levels, the dependence of the response on the bulk lifetime and hence on the generation current within the device, the initial gain, transient response, dependence of the device uniformity upon performance, and the quantity of storable charge within the device. Based on these calculations, we provided insight into possible APD charge storage devices that could be then integrated with an ACT readout device to provide full functionality. In the next section, we discuss our theoretical work on developing the ACT readout device.

C. ACT Readout Architecture

The final component within the ACT imager chip is the ACT readout device itself. After the light has been detected and converted into charge, accumulated and stored for a specified time, it must then be readout for image processing. In most arrays, charge readout is accomplished using a CCD shift register. Increased resolution of a CCD array requires an increased number of elements. As the number of devices increases within a CCD array, the charge transfer efficiency must increase to ensure adequate charge at the output of the array. This places a severe constraint on the quality of each individual device within the CCD array. To overcome this limitation, we proposed to transfer charge using an ACT device. The goal of the theory and modeling part of the program was to design an ACT device for charge transfer and aid the experimental realization of the structure.

In order to handle an ACT device, the modeling tools had to be reconfigured to include the potential created by the surface acoustic wave. In addition, complications due to mixed boundary conditions on the surface needed to be addressed. Specifically, the mixture of Dirichlet and Neumann boundary conditions on a single surface had to be developed. The details of our approach were published in reference [11] which is enclosed at the end of this report.

The modified simulator was then used to study a heterojunction ACT device. The full details of the analysis were published in reference [12] which is enclosed at the end of this report. The structure we envisioned utilizes a p-n junction to deplete the surface layer to support the surface acoustic wave. The usage of the top junction as the channel depletion mechanism makes the device insensitive to the residual surface state density, thus providing a more robust design. The calculated results indicate that an order of magnitude enhancement in charge capacity is possible using this new structure. Transfer efficiency calculations for several different lifetimes in the transport layer show high efficiency values, greater than 79% efficient with a Shockley Read Hall lifetime of 10 ns. These calculations provided the basis for further experimental work.

II. Impact

In addition to the 12 refereed journal papers published by the theoretical group under this program, we have developed an extensive device simulation tool. The full details of its operation have been disclosed in references [13 and 14]. This device simulation tool is the first of its kind to include a nonparabolic energy band structure within a hydrodynamic simulation. Additionally, we have invented some new device

structures. These are the delta doped, quantum well APD device and the pn-np ACT device. Both structures have great potential for advanced device operation.

III. Students Graduated

Ali F. Salem – Ph.D. Spring 1995. Currently at Conexant, Newport Beach, CA.

Joseph Parks, Jr. – Ph.D. Summer 1997. Currently at Intel Corp., Beaverton, Oregon.

Postdoctoral Assistant – Dr. Arlynn Smith, Currently at ITT Optoelectronics, VA.

IV. Future Work

Though much work has been accomplished during this project, there remains much to do. One of the basic problems encountered during the course of this project was the reliability and robustness of each component. It is fairly well known that APDs are notoriously unreliable. Therefore, an array of APDs is not likely to function well. To this end, we have been working at improving the reliability of APD structures. We have recently invented a new procedure by which edge breakdown can be eliminated without the use of guard rings. This procedure we believe will be of great value in developing APD arrays that could then be used to form an imaging chip. Extensive work remains to analyze and develop these new APD structures.

REFERENCES

- ¹ Y. Wang and K. F. Brennan, J. Appl. Phys. **75**, 313 (1994).
- ² Y. Wang and K. F. Brennan, J. Appl. Phys. **76**, 974 (1994).
- ³ W. Quade, E. Scholl, and M. Rudan, Solid State Electron **36**, 1493 (1993).
- ⁴ J. Kolnik, Y. Wang, I. H. Oguzman, and K. F. Brennan, J. Appl. Phys. **76**, 3542 (1994).
- ⁵ Y. Wang and K. F. Brennan, IEEE J. Quantum Electron **30**, 1156 (1994).
- ⁶ K. Brennan, IEEE J. Quantum Electron **QE-22**, 1999 (1986).
- ⁷ A. F. Salem, A. W. Smith and K. F. Brennan, IEEE Trans. Electron Dev. **41**, 1112 (1994).
- ⁸ A. F. Salem and K. F. Brennan, IEEE J. Quantum Electron. **31**, 944 (1995).
- ⁹ A. F. Salem and K. F. Brennan, IEEE Trans. Electron Dev. **43**, 664 (1996).
- ¹⁰ J. W. Parks, Jr., and K. F. Brennan, IEEE Trans. Electron Dev. **45**, 394 (1998).
- ¹¹ A. W. Smith, J. W. Parks, Jr., J. N. Haralson, II, and K. F. Brennan, IEEE Trans. Computer Aided Design of Integrated Circuits and Systems **16**, 420 (1997).
- ¹² A. W. Smith, J. S. Kenney, W. D. Hunt, K. F. Brennan, R. Benz, and C. J. Summers, IEEE Trans. Electron Dev. **42**, 977 (1995).
- ¹³ A. W. Smith and K. F. Brennan, Solid-State Electronics **39**, 1659 (1996).
- ¹⁴ A. W. Smith and K. F. Brennan, Solid-State Electronics **39**, 1055 (1996).

Semiclassical study of the wave vector dependence of the interband impact ionization rate in bulk silicon

Yang Wang and Kevin F. Brennan
*School of Electrical Engineering and Microelectronics Research Center, Georgia Institute of Technology,
Atlanta, Georgia 30332-0250*

(Received 14 July 1993; accepted for publication 28 September 1993)

We present calculations of the interband impact ionization rate calculated using a wave vector dependent (k -dependent) semiclassical formulation of the transition rate. The transition rate is determined using Fermi's golden rule from a two-body screened Coulomb interaction assuming energy and momentum conservation. The transition rate is calculated for the first two conduction bands of silicon by numerically integrating over the full Brillouin zone. The overlap integrals in the expression for the transition rate are determined numerically using a 15 band $k \cdot p$ calculation. It is found that the transition rate depends strongly on the initiating electron wave vector (k vector) and that the transition rate is greatest for electrons originating within the second conduction band than the first conduction band. An ensemble Monte Carlo simulation, which includes the numerically determined ionization transition rate as well as the full details of the first two conduction bands, is used to calculate the total impact ionization rate in bulk silicon. Good agreement with the experimentally determined electron ionization rate data is obtained.

I. INTRODUCTION

The impact ionization process is very important to the study of modern semiconductor devices, such as avalanche photodetectors, in which gain is provided by carrier multiplication. In these devices, the most common impact ionization mechanism present involves an interband transition. For an interband ionization event to occur, the initiating carrier must achieve an energy greater or equal to the energy band gap as measured relative to the band edge, due to energy conservation. Subsequently, a theory of the interband impact ionization process must necessarily address the physics of high energy carrier transport. As is well known, at high carrier energies most of the usual approximations and assumptions, i.e., relaxation time approximation, effective mass approximation, etc. are no longer valid. As a result, it is extremely difficult to fashion a first principles theory of impact ionization since it requires dispensing with virtually all of the standard approaches to semiconductor device modeling and analysis.

The first attempts at formulating a theory of impact ionization were, out of necessity, greatly simplified. Nevertheless, the early theories¹⁻³ provide a useful guide for understanding the underlying physics of the impact ionization process. The theories of Shockley¹ and Wolff² outline opposite and extreme descriptions of the manner by which a carrier achieves sufficiently high energy after which an impact ionization event can occur. This energy is often referred to as the threshold energy. The Shockley theory¹ assumes that the only ionizing carriers are those which are sufficiently "lucky" to drift to an energy greater than or equal to the threshold energy without experiencing any energy relaxing collisions. Conversely, the Wolff theory predicts that ionization events occur after a carrier suffers many collisions while drifting upwards in energy.² Baraff³ later formulated a more complete picture of the impact

ionization process by combining the salient features of the Shockley¹ and Wolff² theories. Though Baraff's theory provides better agreement with realistic systems, it nevertheless fails to predict the impact ionization rate without recourse to the use of adjustable parameters which can not be reliably determined except through comparison to an experiment. As a result, the Baraff theory cannot predict from first principles the impact ionization rate.

In order to progress beyond the limitations imposed by analytical theories, i.e., their reliance on such assumptions as the effective mass and relaxation time approximations, numerical theories of impact ionization have been proposed. These theories⁴⁻⁶ are based on the numerical solution of the Boltzmann equation through the Monte Carlo method and include full details of the band structures as well as the phonon scattering rates at high energies. Although the Monte Carlo method provides a rigorous determination of the distribution function, most present approaches fail to provide a truly first principles calculation of the impact ionization rate owing to their reliance on a parametrized expression for the impact ionization transition rate,⁷ called the Keldysh formula. The use of the Keldysh formula, though universally accepted, is not totally satisfactory since it is derived assuming parabolic energy bands, contains two parameters which must be determined empirically, fails to account for any possible wave vector dependence (k dependence) of the rate, and is determined to only first order in time dependent perturbation theory. Owing to its reliance on empirically determined parameters, the use of the Keldysh formula is limited to studying only those materials for which the electron and hole impact ionization rates have already been experimentally determined. Therefore, a simulator which incorporates the Keldysh formula is of limited usefulness when investigating the behavior of materials and devices comprised of technologically immature materials systems for

which no reliable experimental data of the ionization rates are available.

The first attempt to improve the impact ionization formulation was provided by Kane.⁸ In Kane's formulation the pair cross section for the impact ionization transition rate in bulk silicon is calculated using pseudopotential generated energy bands and wave functions. Kane compared the transition rate calculated assuming momentum conservation as well as a random k approximation in which momentum conservation is effectively ignored. Interestingly, he found reasonable agreement in the calculated transition rates averaged over energy in either approach. In other words, if the transition rate is determined by averaging over a suitable set of wave vectors (k vectors or k states) for a given energy then the k dependence is effectively averaged out. Nevertheless, the transition rate at any one initial k state may be quite different from that at another k state of comparable energy. More recently, Sano *et al.*^{9,10} and Bude and Hess¹¹ have investigated the threshold energy in silicon and GaAs. Their work indicates that the threshold is highly k dependent and that the relative softness of the threshold originates from the k dependence of the impact ionization transition rate.

New theories have examined the impact ionization transition rate semiclassically¹¹ and with a more complete quantum mechanical picture.¹²⁻¹⁴ These works have greatly improved the understanding of impact ionization by demonstrating that the threshold energy is always relatively soft owing to energy broadening effects from the electron-phonon interaction. Bude and Hess¹¹ have shown, using a Kane-like calculation, that the ionization rate does not necessarily truncate the high energy tail until very high energies are attained. In other words, some of the electrons can survive to relatively high energies before undergoing impact ionization, in contrast to a hard threshold model.

In this article, we calculate the impact ionization rate in bulk silicon using a semiclassical model much like that employed by Kane⁸ and Bude and Hess¹¹ but do not average the transition rate over the energy. Instead, we calculate the electron impact ionization transition rate at different k points for a mesh incorporated within the first Brillouin zone. Our approach is similar to that employed by Yoder *et al.*¹⁴ In their work, Yoder *et al.*¹⁴ further include a formulation of the phonon scattering rate with the same pseudopotentials used for calculating the band structure. However, Yoder *et al.*¹⁴ do not report either the ionization transition rate or the overall impact ionization rate calculated using their model, only the average electron energy as a function of electric field. As such it is difficult to compare to their theory. The k -dependent impact ionization transition rate mesh is included within an ensemble Monte Carlo simulator^{15,16} which contains the full details of the first two conduction bands in silicon. The calculations are compared to both experimental measurements as well as other theoretical calculations. The details of the transition rate calculation are reviewed in Sec. II along with the salient features of the Monte Carlo model. Calculated results are presented in Sec. III and conclusions are drawn in Sec. IV.

TABLE I. Material parameters for bulk silicon simulation (top) and intervalley coupling constants and phonon energies (bottom).^a

Parameter	Value and units
Density ^a	2.329 g/cm ³
Lattice constant ^a	5.43 Å
Energy gap ^a	1.12 eV
Dielectric constant (relative) ^a	11.7
Effective masses (relative)	
<i>L</i> valley ^a	0.284
<i>X</i> valley ^a	0.321
Heavy hole ^b	0.537
Light hole ^b	0.153
Split-off hole ^b	0.234
Valley separation energies ^c	
Δ _{XL}	0.8 eV
Δ _{XT}	2.3 eV
Acoustic deformation potential ^d	9.0 eV
Optical phonon energy ^d	0.063 eV
Sound velocity ^d	
Longitudinal	9.04 × 10 ⁵ cm/s
Transverse	5.34 × 10 ⁵ cm/s

	Phonon temperature (K)	Deformation potential (eV/cm)	Scattering type
<i>X-X</i>	220	3 × 10 ⁷	f
	550	2 × 10 ⁸	f
	685	2 × 10 ⁸	f
	140	5 × 10 ⁷	g
	215	8 × 10 ⁷	g
	720	1.1 × 10 ⁹	g
<i>X-L</i>	672	2 × 10 ⁸	
	634	2 × 10 ⁸	
	480	2 × 10 ⁸	
	197	2 × 10 ⁸	

^aReference 16.

^bReference 17.

^cReference 20.

^dReference 18.

II. MODEL DESCRIPTION

The total impact ionization rate is calculated by including a numerically determined semiclassical impact ionization transition rate into an ensemble Monte Carlo model. The carrier histories within the Monte Carlo simulator are traced in three dimensions in both real and k space. The details of the Monte Carlo model employed in this study have been exhaustively presented elsewhere.^{15,16} The full details of the band structure are included in the Monte Carlo model.^{15,16} The method by which phonon scattering is accounted for is described below. The parameters used for the calculation of the scattering rate in silicon are presented in Table I and have been taken from Refs. 17–19 as marked in Table I.

The important feature employed within the present model is the use of a k -dependent impact ionization transition rate. The impact ionization process is treated as an additional scattering mechanism in the Monte Carlo program. The manner in which the transition rate is calculated can be summarized as follows. The transition rate is calculated assuming that a target electron is always available in the valence band, the final states are always empty (non-

degenerate conditions) and Fermi's golden rule applies as,²⁰

$$W_{\text{imp}} = \int \frac{2\pi}{\hbar} |M|^2 \delta(E_f - E_i) dS_f, \quad (1)$$

where dS_f , as in Ref. 20, represents integration over all the final states. The matrix element, M , is taken to be a screened Coulomb interaction between two electrons and can be expressed as

$$M = \frac{e^2}{\epsilon V} \frac{I(k_1, k'_1) I(k_2, k'_2)}{|k_1 - k'_1|^2 + q_0^2}, \quad (2)$$

where k_1 and k'_1 are the k vectors corresponding to the incident electron both before and after the interaction, respectively, and k_2 and k'_2 represent the target electron k vectors both before and after the collision. The quantities $I(k_1, k'_1)$ and $I(k_2, k'_2)$, are overlap integrals and q_0 is the static screening factor. Following the approach of Ridley,²⁰ we assume q_0 to be the inverse Debye length.

The transition rate is evaluated numerically by integrating Eq. (1) with Eq. (2) over the full Brillouin zone. The overlap integrals used in the calculation are determined by diagonalizing a 15×15 $k \cdot p$ Hamiltonian.²¹ The parameters used within our $k \cdot p$ calculation are the same as those reported in Ref. 21. The transition rate needed here requires that its absolute magnitude be determined. As such, the delta function must be properly normalized by ensuring that its integration alone always yields one. The overlap integrals in the matrix element are determined using $k \cdot p$ generated wave functions. The values of the overlap integrals obtained are in excellent agreement with those determined by Burt *et al.*²² The transition rate is determined for each mesh point within a finely spaced, 916 point k space grid spanning the reduced zone⁴ of the first Brillouin zone which is consistent with the grid spacing used in Ref. 5 for determining the energy. Additional points are also used for interpolation purposes raising the total number of points at which the transition rate is calculated to 1419. The transition rate is calculated for both the first and second conduction bands at each mesh point. Several million final k points are sampled in order to ensure good convergence in the integration. The numerically generated grid is used to determine the impact ionization transition rate for the electrons as they drift through the crystal. During the course of the simulation, after each drift, the electron's energy is first checked to see if it exceeds or equals the energy gap. If the electron energy is less than the energy gap, it cannot undergo an impact ionization event and the ionization transition rate is simply zero for that electron. The probability that the electron scatters is then found in the usual manner.^{4,5} When an electron has energy equal to or greater than the energy gap, the ionization transition rate is determined from use of the k space grid. The electron is located by its k vector within some subvolume of the k space grid. If the ionization transition rate is nonzero for all of the mesh points surrounding the given k vector then the transition rate for that particular electron is determined by interpolating linearly between

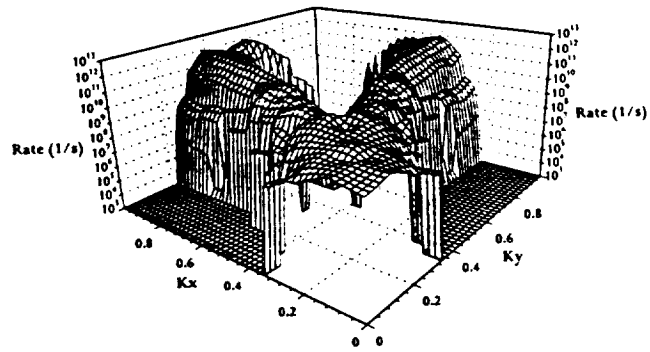


FIG. 1. Calculated interband impact ionization transition rate within the first conduction band in bulk silicon as a function of k_x and k_y at fixed k_z . k_z is held fixed at 0.0. Those points in the diagram which have a transition rate corresponding to the minimum of the y axis scale have no threshold and hence have a zero rate. Owing to the logarithm scale used in the figure, we have chosen to represent these points as having a minimum transition rate for convenience.

the values at each mesh point. If one or more of the mesh point values of the transition rate is zero, then the transition rate for the desired k vector is chosen equal to that of the nearest k point. The impact ionization transition rate determined in this manner is then added to the total phonon scattering rate corresponding to the energy of the electron. As in standard Monte Carlo algorithms, a random number is thrown to determine if the carrier is scattered. Given that a carrier has scattered, another random number is chosen to determine the particular event that occurs out of all of the possible events, phonon scatterings, and impact ionization.

The calculated interband impact ionization transition rate for the first and second conduction bands of silicon as a function of k_x and k_y , holding k_z fixed at 0.0 is presented in Figs. 1 and 2. As can be seen from these figures, the transition rate is substantially higher for electrons originally confined within the second conduction band than the first conduction band. The maximum transition rate overall is larger than 10^{13} 1/s but less than 10^{14} 1/s for the k

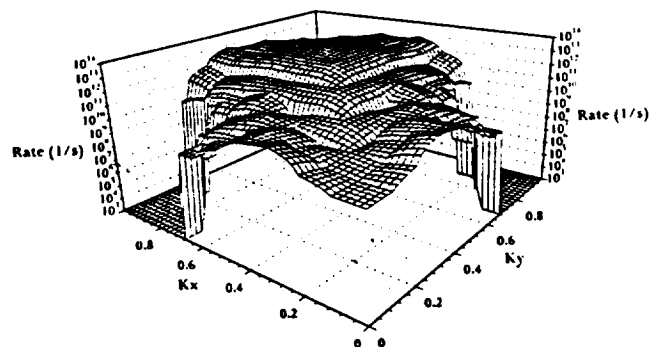


FIG. 2. Calculated interband impact ionization transition rate within the second conduction band in bulk silicon as a function of k_x and k_y at fixed k_z . k_z is held fixed at 0.0. Those points in the diagram which have a transition rate corresponding to the minimum of the y axis scale have no threshold and hence have a zero rate. Owing to the logarithm scale used in the figure, we have chosen to represent these points as having a minimum transition rate for convenience.

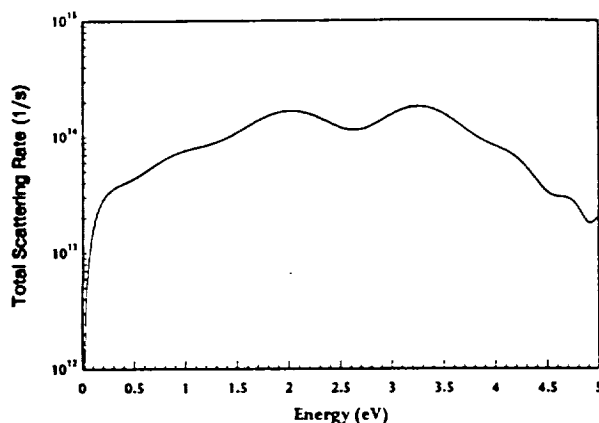


FIG. 3. Calculated electron-phonon scattering rate in bulk silicon as a function of electron energy. The scattering rate is calculated including the first two conduction bands. The low energy scattering rate is determined using Fermi's golden rule and the parameters listed in Table I. The high energy scattering rate is directly proportional to the final density of states. The two peaks which appear in the diagram are due to peaks in the density of states of the first two conduction bands.

points considered here. Notice that the transition rate for electrons confined within the first conduction band does not exceed 10^{13} 1/s and is typically much less than 10^{13} 1/s. This is apparently in agreement with Bude and Hess¹¹ who found that the transition rate is substantially lower than the corresponding phonon scattering rate in general.

During the course of the Monte Carlo simulation, the electrons can move between the two conduction bands only as a result of a scattering event; interband tunneling transitions are not included in the simulation. The manner by which the carriers change bands can be explained as follows. An array of possible k states in each band, ordered by increasing energy, is first generated. Each array element contains the k vector, energy, and band index of the state. After a phonon scattering event, the final state is selected stochastically from the sorted array in accordance with the physics of the particular scattering mechanism present. Typically, at high electron energies, the phonon scattering events are predominately deformation potential scatterings. For simplicity, it is assumed here that the deformation potential scatterings have no angular dependence which is the standard assumption. More refined theories of deformation potential scattering¹⁴ indicate that they might show some angular dependence. The high energy phonon scattering rates, in our calculation, are taken then to be proportional to the total density of final states, including both conduction bands.¹¹ Therefore, an electron originally confined within band 1 can be scattered into band 2 if band 2 overlaps band 1 at an energy corresponding to the final state energy. In this way, interband transitions can be automatically accounted for in the simulation.

The total phonon scattering rate as a function of electron energy is plotted in Fig. 3. The low energy scattering rate is calculated from Fermi's golden rule using the parameters specified in Table I. The scattering mechanisms included and their formulations were taken from Ref. 23. At high carrier energies, the phonon scattering rate is as-

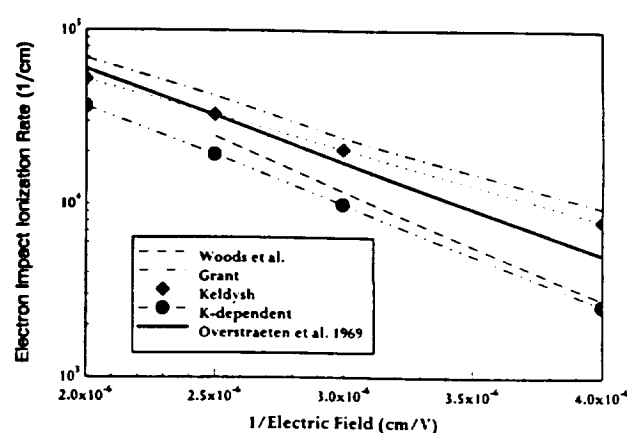


FIG. 4. Calculated and experimental results for the electron impact ionization rate in bulk silicon plotted as a function of inverse electric field. Calculations of the ionization rate including the Keldysh formula as well as the k -dependent model are presented.

sumed, as stated above, to be dominated by deformation potential scattering. The scattering rate at these energies is determined by directly integrating over the final density of states calculated from the numerically generated band structure including collision broadening. Collision broadening of the final state is treated following the approach of Ref. 24. A single deformation potential constant is selected in order to match the scattering rate calculated from Fermi's golden rule at 0.4 eV. Subsequent calculations of the steady-state drift velocity show good agreement with the experimental saturation carrier drift velocity in bulk silicon.²⁵ Inspection of Fig. 3 shows that the phonon scattering rate exceeds 10^{14} 1/s. Comparison of the phonon scattering rate and the impact ionization transition rate shows that for most points within the reduced zone, the phonon scattering rate is substantially higher than the impact ionization rate.

III. CALCULATED RESULTS

The calculated impact ionization rate using the k dependent formulation as a function of inverse applied electric field is plotted in Fig. 4 along with the experimental measurements of Overstraeten *et al.*,²⁶ Grant,²⁷ and Woods *et al.*²⁸ Although other experimental data exist, these data comprise a representative set and bracket the range of experimental measurements. As can be seen from Fig. 4, the k dependent calculations match best to the low range of the experimental data, those of Woods *et al.*,²⁸ throughout the range of applied electric field strengths considered here. Additional calculations using the Keldysh formulation for the impact ionization transition rate within the otherwise identical Monte Carlo model are included for comparison. The calculations using the Keldysh formula, with the particular choice of p and E_{th} chosen here, show better agreement with the higher range of experimental measurements, those corresponding to Grant,²⁷ particularly at lower electric field strengths. The choice of p and E_{th} used in the Keldysh formula are 0.01 and 1.13 eV, respectively. The value of 0.01 for p is in the range of values which typically

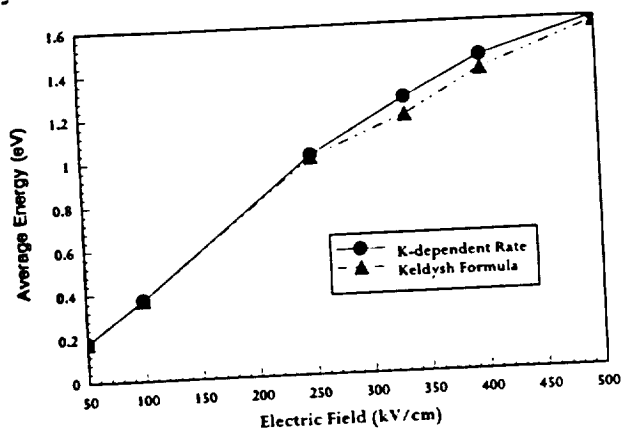


FIG. 5. Calculated average electron energy as a function of applied electric field within the k -dependent and Keldysh formula models. At low electric fields, both models predict exactly the same average energy as expected since no ionization events occur and there is no difference between the two methods. As the field strength increases, the average energy is somewhat lower in the Keldysh formula model than in the k -dependent model. This is due to the fact that in the Keldysh model the transition rate is precisely the same for carriers of the same energy independent of their k vector or band index. Subsequently, more electrons ionize from within the lower energy, first conduction band than from the second conduction band. In contrast, within the k -dependent model, the electron ionization rate is significantly greater for electrons within the second conduction band than the first conduction band. Hence, the carriers survive to higher energies before ionizing. The effect is not dramatic however, since the relative number of ionization events is not large. Nevertheless, some of the overall trend is apparent.

correspond to a soft threshold condition. It should be noted that the Keldysh formula calculations could be fitted to the lower ionization rate data just as easily by changing the parameters p and E_{th} . Little new information about the nature of the threshold, i.e., softness or hardness, would be obtained by doing this however since the parameters would change relatively very little. Therefore, we have chosen the results reported here just to illustrate the order of magnitude of p that the Keldysh formula predicts within this specific Monte Carlo model.

The average energy of the electrons as a function of applied electric field for both models is plotted in Fig. 5. As can be seen from the figure, the average energy of the electrons is quite high increasing to almost 1.6 eV at an applied field of 500 kV/cm. This implies that the threshold is exceedingly soft, in that many of the electrons survive to energies well above the band gap energy before impact ionizing, in good agreement with the recent work of Bude and Hess.¹¹ As discussed by Bude and Hess,¹¹ the relatively weak ionization transition rate fails to truncate the electron distribution function at energies near the band gap (as would be the case in a hard threshold process) leading to very high average carrier temperatures.

We have also investigated the effect of the second conduction band on the impact ionization rate in bulk silicon. The percentage of ionization events per band at various applied electric field strengths is plotted in Figs. 6 and 7 for the k dependent and Keldysh formula models, respectively. Inspection of Fig. 6 clearly shows that most of the impact ionization events originate from the second conduction

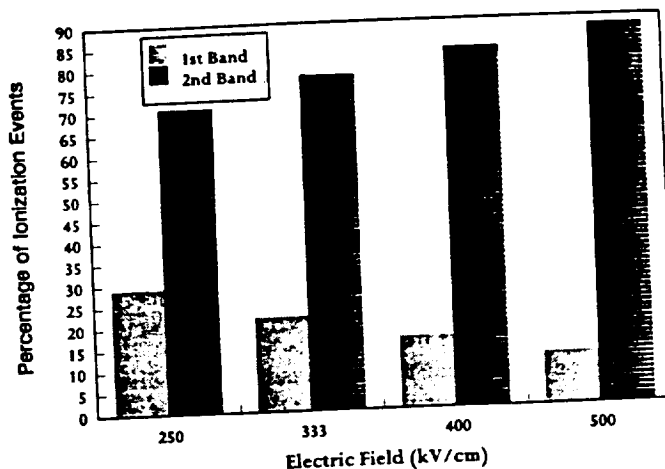


FIG. 6. Histogram showing the percentage of electron ionization events originating from the first and second conduction bands as a function of applied electric field within the k -dependent model. Notice that the vast majority of events originate from the second conduction band even at low applied electric fields.

band in the k -dependent model even at low applied electric field strengths. This is apparently due to the fact that the transition rate is significantly higher in the second band than the first band as is seen from Figs. 1 and 2. Therefore, in the k -dependent model, most of the ionization events occur only after an electron has transferred into the second conduction band which occurs predominately at high electron energies. As a result, the average energy of the electrons is necessarily quite high. As the electric field strength increases, the percentage of events within the second band increases as well. This behavior is as expected since at higher electric fields and hence, higher carrier energies the electrons will transfer more readily to the second conduction band.

Figure 7 shows the percentage of electron ionization events originating from the first and second conduction

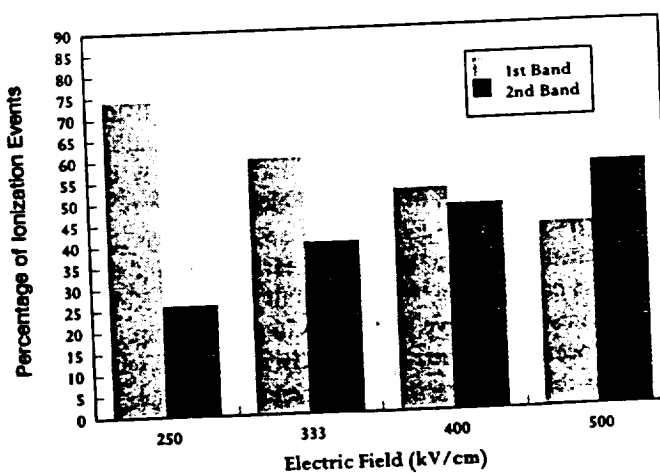


FIG. 7. Histogram showing the percentage of electron ionization events originating from the first and second conduction bands as a function of applied electric field based on the Keldysh formula model. At low applied electric fields, most of the ionization events originate from the first conduction band in contrast to the k -dependent model.

bands in the Keldysh formula model. In contrast to the k -dependent model, most of the ionization events originate from the first conduction band at low applied electric field strengths. In the Keldysh formula model, all states of the same energy have precisely the same ionization transition rate and no distinction is made between states in the first and second conduction bands. The electrons have the same probability at a given energy of ionizing from the first conduction band as from the second conduction band in the Keldysh model while in the k -dependent model, impact ionization is far more likely for electrons within the second conduction band than the first even for the same energy. Therefore, it is not surprising that the Keldysh model predicts that more events will occur from the first conduction band than the second at lower fields. Additionally, one would expect that the average electron energies would be predicted to be smaller based on the Keldysh formula model than the k -dependent model once impact ionization has begun. Inspection of Fig. 5 shows that the average electron energy between the two models is the same at low applied electric field strengths where no ionization occurs but it is somewhat lower in the Keldysh model than the k -dependent model at higher electric fields. The observed difference at high fields is not dramatic due to the fact that the number of ionization events is relatively small overall. Hence, the ionization rate does not drastically change the average electron energies. Nevertheless, some overall trend is present.

An energy dependent rate can be derived from the k -dependent rate in a manner analogous to that done by Kane.⁸ An additional integration over the initiating electron k -vector space, S_i , is performed following Kane⁸ as

$$W(E) = \sum_n \left(\frac{1}{2\pi} \right)^3 \int dS_i W_{\text{imp}}(n, k) \delta[E_i(n, k) - E], \quad (3)$$

where the sum is performed over the first two conduction bands, n is the band index, and E_i is the initiating electron energy. The value for $W(E)$ needs, in addition, to be normalized by the density of states at the energy E as was done by Kane.⁸ This has been done here as well. It is interesting to compare $W(E)$ to the energy-dependent rate specified by the Keldysh formula. $W(E)$ and the Keldysh rate are plotted as a function of initiating electron energy in Fig. 8. The curve for $W(E)$ in Fig. 8 has been purposely "smoothed" in order to more clearly show its shape and facilitate comparison with the result from the Keldysh formula. As stated above, from Figs. 6 and 7, more ionization events occur from carriers initially within the second conduction band using the k -dependent model than the Keldysh model. Therefore, one would expect that the transition rate would be greater at higher energy for $W(E)$ than for the Keldysh formula. This is clearly indicated in Fig. 8; the impact ionization transition rate calculated by averaging over the k -dependent rate, $W(E)$, is higher at higher energy and lower at lower energy than that calculated using the Keldysh formula.

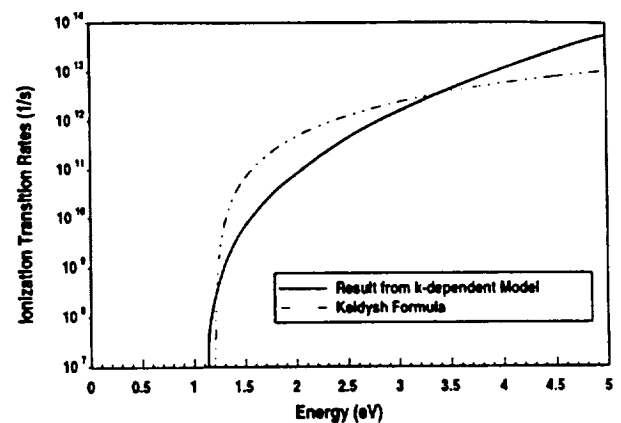


FIG. 8. Calculated impact ionization transition rates as a function of electron energy measured from the conduction band minimum using the k -dependent and Keldysh models. The energy dependent transition rate is determined by performing a further integration of the k -dependent rate as described in the text.

IV. CONCLUSIONS

In this article we have presented calculations of the electron interband impact ionization rate calculated using a k -dependent semiclassical formulation of the transition rate. Fermi's golden rule is used to determine the transition rate from a two-body screened Coulomb interaction assuming energy and momentum conservation. The transition rate is calculated by numerically integrating over several million final states distributed within the first Brillouin zone. The transition rate is determined at each of 916 mesh points within the reduced zone forming a finely spaced k -space grid spanning both the first two conduction bands of silicon. It is found that the transition rate depends strongly on the initiating electron k vector and that the transition rate is greater for electrons originating within the second conduction band than the first conduction band in bulk silicon. An ensemble Monte Carlo simulation, which includes the numerically determined ionization transition rate as well as the full details of the first two conduction bands, is used to calculate the total impact ionization rate in bulk silicon. Good agreement between the k -dependent model and the experimental measurements of Woods *et al.*²⁸ is attained.

The important result of this work is that the calculations based on the k -dependent model show that the ionization rate in silicon is best modeled as a soft threshold process. Owing to the fact that the transition rate is significantly lower than the phonon scattering rate and that the transition rate is greatest within the second conduction band, the electrons drift to high energies, on average, before undergoing an impact ionization event. As a result, the average energy of the electrons is quite high at electric field strengths for which impact ionization can occur. The ionization rate fails to truncate the electron distribution at energies near the band gap or threshold energy. These results are qualitatively in excellent agreement with those reported by Bude and Hess.¹¹ Recent experimental work of Cartier *et al.*²⁹ using the X-ray photoemission spectroscopy

(XPS) technique indicates that the ionization rate may have multiple ranges of behavior and that the higher range experimental data of Grant²⁷ may be most accurate. Although our model predicts a significantly lower ionization rate than that of Grant,²⁷ we believe that this may be due principally to the failure of the semiclassical approach. A more complete theory including quantum mechanical effects, details of the scattering rates, and band structures at high energies is probably a more appropriate approach for predicting impact ionization rates. We are currently pursuing such a model.

ACKNOWLEDGMENTS

This work was sponsored in part by ARPA and NASA through contract to NASA, NAGW-2753. Additional support was received by the National Science Foundation through a Presidential Young Investigator Award made to K. Brennan. Computing support was provided by the Digital Equipment Corporation through contract E21-H36.

¹W. Shockley, *Solid-State Electron.* 2, 35 (1961).

²P. A. Wolff, *Phys. Rev.* 95, 1415 (1954).

³G. A. Baraff, *Phys. Rev.* 128, 2507 (1962).

⁴H. Shichijo and K. Hess, *Phys. Rev. B* 23, 4197 (1981).

⁵M. V. Fischetti, *IEEE Trans. Electron. Dev.* 38, 634 (1991).

⁶For an in-depth review of the theories of impact ionization see F. Capasso, "Physics of Avalanche Photodiodes," in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beer (Academic, New York, 1985), Vol. 22, Part D.

⁷L. V. Keldysh, *Zh. Eksp. Teor. Fiz.* 48, 1962 (1965) [*Sov. Phys. JETP* 21, 1135 (1965)].

⁸E. O. Kane, *Phys. Rev.* 159, 624 (1967).

⁹N. Sano, T. Aoki, and A. Yoshii, *Appl. Phys. Lett.* 55, 1418 (1989).

¹⁰N. Sano, M. Tomizawa, and A. Yoshii, *Appl. Phys. Lett.* 56, 653 (1990).

¹¹J. Bude and K. Hess, *J. Appl. Phys.* 72, 3554 (1992).

¹²J. Bude, K. Hess, and G. J. Iafrate, *Phys. Rev. B* 45, 10958 (1992).

¹³W. Quade, F. Rossi, and C. Jacoboni, *Semicond. Sci. Technol.* 7, B502 (1992).

¹⁴P. D. Yoder, J. M. Higman, J. Bude, and K. Hess, *Semicond. Sci. Technol.* 7, B357 (1992).

¹⁵K. F. Brennan, D. H. Park, and Y. Wang, *IEEE Trans. Electron. Dev.* 37, 536 (1990).

¹⁶K. F. Brennan, N. Mansour, and Y. Wang, *Computer Phys. Comm.* 67, 73 (1991).

¹⁷J. Y. Tang, Ph.D. thesis, University of Illinois, Urbana, IL, 1982.

¹⁸Y. Ohno, *J. Appl. Phys.* 64, 4549 (1988).

¹⁹C. Jacoboni, C. Canali, G. Ottaviani, and A. A. Quaranta, *Solid-State Electron.* 20, 77 (1977).

²⁰B. K. Ridley, *Quantum Processes in Semiconductors*, 2nd ed. (Oxford University Press, Oxford, 1988).

²¹F. H. Pollak, C. W. Higginbotham, and M. Cardona, *J. Phys. Soc. Jpn.* 21, 20 (1966).

²²M. G. Burt, S. Brand, C. Smith, and R. A. Abram, *J. Phys. C: Solid State Phys.* 17, 6385 (1984).

²³C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer, Wien, 1989).

²⁴Y. C. Chang, D. Z.-Y. Ting, J. Y. Tang, and K. Hess, *Appl. Phys. Lett.* 42, 26 (1983).

²⁵C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, and A. Alberigi-Quaranta, *Phys. Rev. B* 12, 2265 (1975).

²⁶R. van Overstraeten and H. DeMan, *Solid-State Electron* 13, 583 (1970).

²⁷W. N. Grant, *Solid-State Electron.* 16, 1189 (1973).

²⁸M. H. Woods, W. C. Johnson, and M. A. Lambert, *Solid-State Electron.* 16, 381 (1973).

²⁹E. Cartier, M. V. Fischetti, E. A. Eklund, and F. R. McFeely, *Appl. Phys. Lett.* 62, 3339 (1993).

Numerical study of the wave-vector dependence of the electron interband impact ionization rate in bulk GaAs

Yang Wang and Kevin F. Brennan

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0250

(Received 16 December 1993; accepted for publication 4 April 1994)

Ensemble Monte Carlo calculations of the electron interband impact ionization rate in bulk GaAs are presented using a wave-vector (k)-dependent formulation of the ionization transition rate. The transition rate is evaluated through use of numerically generated wavefunctions determined via a k - p calculation within the first two conduction bands at numerous points within a finely spaced three-dimensional grid in k space. The transition rate is determined to be greatest for states within the second conduction band. It is found that the interband impact ionization transition rate in bulk GaAs is best characterized as having an exceedingly "soft" threshold energy. As a consequence, the dead space, defined as the distance over which the ionization probability for a given carrier is assumed to be zero, is estimated to be much larger than that estimated using a "harder" threshold. These results have importance in the design of multiquantum-well avalanche photodiodes.

1. INTRODUCTION

The probability that an interband impact ionization event will occur depends upon the likelihood that a carrier will first drift to an energy equal to or greater than the band-gap energy, often referred to as the threshold energy, and then suffer an impact ionization event before phonon scattering processes relax it back to an energy below threshold. Most of the early analytical theories of interband impact ionization^{1,2} concerned themselves with determining the probability that a carrier survives to threshold assuming that it would ionize immediately afterward. Such theories are said to assume a "hard threshold" for interband impact ionization.

The hard threshold model was not limited to the early analytical theories. More advanced theories^{3,4} were formulated using a hard threshold model for interband impact ionization; however, more recent numerical studies^{5,6} as well as analytical theories^{7,8} have called this assumption into question. Interestingly, in most of the numerical theories employed in the past the calculated impact ionization rates often agree well with experimental results for some semiconductor materials using either a hard or "soft" threshold model.⁹ This is due to the fact that most numerical formulations rely on a parameterized expression, called the Keldysh formula,¹⁰ to describe the impact ionization transition rate. The Keldysh formula contains two parameters, P and E_{th} , which are very difficult to determine from first principles. E_{th} corresponds to the threshold energy for impact ionization while P is some measure of how strong the impact ionization transition rate is for carriers with energies greater than E_{th} . A small value of P then corresponds to a soft threshold since the transition rate is relatively small near the threshold and increases only slowly thereafter. In contrast, a hard threshold is characterized by a large value of P . Typically, both P and E_{th} are selected such that the calculated impact ionization rate is in agreement with experimental measurements. As such, more than one set of parameters can often reproduce the same ionization rate in a given material. Tang and Hess⁹ have shown that the calculated impact ionization rate in bulk sili-

con, based on an ensemble Monte Carlo model including the Keldysh formula, agrees well with existing experimental data using either a soft or hard threshold model.

Kane¹¹ made the first attempt to improve the impact ionization transition rate formulation by directly evaluating the pair cross section using pseudopotential-generated energy bands and wave functions for bulk silicon. In his formulation, Kane calculated the transition rate for different wave vector (k vector) states of comparable energy and averaged these rates to determine an energy dependent transition rate. More recently, Sano *et al.*^{12,13} and Bude and Hess⁵ have investigated the threshold energy in silicon and GaAs using numerical techniques. Their work indicates that the threshold is highly k dependent and that the relative softness of the threshold originates from the k dependence of the impact ionization transition rate.

New theories of interband impact ionization have been presented which treat the problem either semiclassically^{5,14} or quantum mechanically.¹⁵⁻¹⁷ The quantum-mechanics works have demonstrated that the threshold energy is always relatively soft owing to energy broadening effects arising from the electron-phonon interaction. In the quantum mechanical formulations, a Lorentzian is substituted in place of the energy conserving delta function which leads to broadening of the threshold state itself. Bude and co-workers¹⁵ have demonstrated that through inclusion of quantum-mechanical broadening, a threshold state cannot rigorously be defined since the Lorentzian does not demand strict energy conservation over short time scales. As a result, a carrier can impact ionize even if its energy is less than the energy gap. Energy broadening acts then to inflate the impact ionization rate near threshold.

Similarly, even the semiclassical models predict that the ionization transition rate in bulk silicon is relatively low leading to a soft thresholdlike behavior. Owing to the fact that the ionization transition rate is calculated to be significantly lower than the competing phonon scattering rate, these models predict that the electrons drift to high energies, on average, before suffering an impact ionization event. As

result, the average energy of the electrons is expected to be quite high at high-electric-field strengths in contrast to what is predicted using a hard threshold model. Recent experimental measurements using x-ray photoemission spectroscopy (XPS)¹⁸ indicate that the threshold for interband impact ionization in bulk silicon is exceedingly soft. These results further indicate that the ionization rate may have multiple ranges of behavior which is consistent with theoretical predictions^{5,14} that the second band dominates the impact ionization process in bulk silicon.

Aside from the obvious importance of improving the understanding of interband impact ionization, knowledge of the nature of the threshold energy in different materials is useful in modeling advanced avalanche photodiode (APD) designs which utilize heterostructures and multiple quantum wells to enhance the electron ionization rate.¹⁹⁻²² In these devices the presence of the conduction-band-edge discontinuity formed at the heterostructure interface provides a kinetic energy boost to electrons injected into the narrow-band-gap layers from the wide-band-gap materials. The larger the conduction-band-edge discontinuity is, the greater relative effect it has on the electron ionization rate. This depends though upon how soft the threshold is. If the threshold is relatively hard, the conduction-band-edge discontinuity can greatly influence the ionization rate by providing sufficient energy to promote the electron from below threshold to above threshold. On the other hand, if the threshold is exceedingly soft, the presence of the conduction-band-edge discontinuity may not have as much of an effect since most of the carriers must attain very high energies before impact ionizing. As a result, the discontinuity may contribute relatively little to the overall heating of the electron distribution.

Since interband impact ionization is a threshold process, a carrier launched near the band edge, either initially or after suffering an ionization event, must drift for some time to acquire sufficient kinetic energy in order to impact ionize. The ionization probability then is necessarily equal to zero for some distance for both the generated carrier and the initiating carrier immediately following an ionization event. The distance over which the ionization probability is assumed to be zero following launching from the band edge is often referred to as the dead space. The dead space can be of importance in multiquantum-well structures since if the well widths are much smaller than the dead space distance, then the ionization rate may be appreciably reduced, leading to little enhancement. Recently, the electron and hole ionization coefficients were measured in a series of GaAs/AlGaAs multiple quantum wells with varying well and barrier widths.²³ It was found that in the narrow well width structures (<100 Å) there was virtually no enhancement of the experimentally measured electron to hole ionization rates ratio, while in wider well structures there was a significant increase in the ionization rates ratio. These results may be related to the dead space in the ionization process.

In this article we present detailed calculations of the electron-initiated impact ionization rate in bulk GaAs using a semiclassical formulation of the interband impact ionization transition rate. The transition rate is determined using Fermi's golden rule from a two-body screened Coulomb inter-

action assuming both energy and momentum conservation for the first two conduction bands of GaAs. These results are then incorporated into an ensemble Monte Carlo simulator to calculate the total impact ionization rate as a function of applied electric field in bulk GaAs. Aside from the ionization transition rate and the total impact ionization rate, the dead space for electron-initiated impact ionization is also calculated and compared to results based on a relatively hard threshold model. The implications of an exceedingly soft threshold for impact ionization in GaAs are discussed in relation to multiquantum-well devices. The details of the model are summarized in Sec. II. The calculated results are presented in Sec. III while conclusions are drawn in Sec. IV.

II. MODEL DESCRIPTION

The details of the model have been exhaustively reviewed in a previous work.¹⁴ Nevertheless, for completeness, the salient features of our model are briefly outlined here. The total impact ionization rate is determined using an ensemble Monte Carlo simulation which includes the full details of the first two conduction bands of GaAs and the numerically generated semiclassical impact ionization transition rate. The Monte Carlo simulator used in the calculations is based on the original work of Shichijo and Hess,²⁴ with the improved phonon scattering technique of Chang *et al.*²⁵ and the improved band-structure interpolation technique of Fischetti and Laux.²⁶ The material parameters used in the simulation are taken from Refs. 27 and 28. The estimators used in the simulation are discussed at length in Ref. 28.

The high-energy scattering rate is assumed to be dominated by deformation potential scattering. The deformation potential scattering rate is calculated using a full order time-dependent perturbation theory expansion assuming that the deformation potential is a constant, independent of energy.²⁵ Although the assumption of a constant deformation potential may not be strictly valid,¹⁷ it is standard and commonly used. The deformation potential scattering is determined by direct integration over the numerically generated density of states for both bands. At the points in k space at which both bands overlap, the deformation potential scattering depends, of course, on the density of states within both bands. During the course of the Monte Carlo simulation, the electrons can transfer then from one band to the other through the action of the deformation potential scattering.

Impact ionization is incorporated into the Monte Carlo model using the semiclassical, wave-vector (k)-dependent transition rate given by Ridley.²⁹ The matrix element is assumed to be a screened Coulomb interaction between two electrons.¹⁴ The transition rate is evaluated numerically by integrating over the full Brillouin zone. Several million possible final states are sampled for each integration to ensure convergence and accuracy. The overlap integrals present in the matrix element are determined by diagonalizing a 15×15 $k \cdot p$ Hamiltonian.³⁰ The values of the overlap integrals match closely to those presented by Burt *et al.*³¹ The transition rate is determined for each mesh point within a finely spaced, 916 point, k -space grid spanning the reduced zone²⁴ of the first Brillouin zone. Additional points are also used for interpolation purposes raising the total number of mesh points at

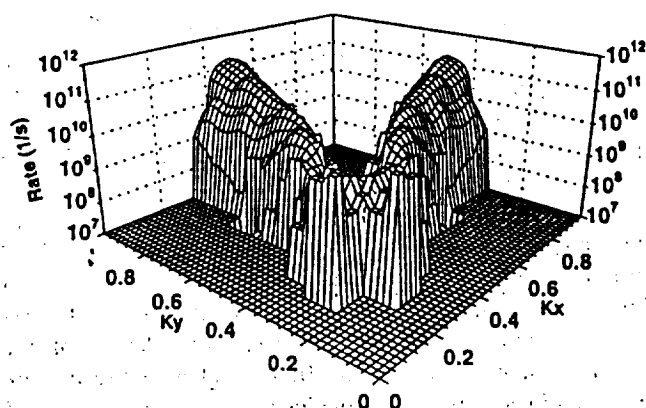


FIG. 1. Calculated interband impact ionization transition rate for the first conduction band of bulk GaAs within the reduced zone of the first Brillouin zone. The ionization transition rate is plotted vs k_x and k_y , holding k_z fixed at 0.0. Those points which are shown to have a transition rate equal to the minimum of the vertical axis scale have no threshold and hence have a zero transition rate. Owing to the logarithmic scale used in the figure, these points are represented as having a minimum transition rate for convenience.

which the transition rate is calculated to 1419. The transition rate is determined for both conduction bands at each of the mesh points. The mesh employed for the impact ionization transition rate is the same as that for the energy interpolation used during the course of the Monte Carlo simulation.²⁶

The instantaneous electron-impact ionization transition rate is determined both from its energy and \mathbf{k} vector. When the simulated electron's energy is less than the energy band gap, the electron cannot impact ionize and its flight is determined in the usual manner.³² When the simulated electron drifts to an energy greater than or equal to the energy band gap, its ionization transition rate is determined using the \mathbf{k} space mesh. The eight mesh points nearest to the components of the simulated electron's \mathbf{k} vector are first determined. If the ionization transition rate is nonzero for all of the eight mesh points then the transition rate of the simulated electron

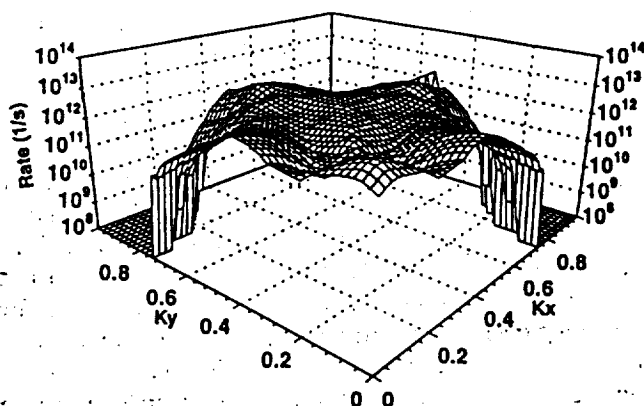


FIG. 2. Calculated interband impact ionization transition rate for the second conduction band of bulk GaAs within the reduced zone of the first Brillouin zone. The ionization transition rate is plotted vs k_x and k_y , holding k_z fixed at 0.0. Those points which are shown to have a transition rate equal to the minimum of the vertical axis scale have no threshold and hence have a zero transition rate. Owing to the logarithmic scale used in the figure, these points are represented as having a minimum transition rate for convenience.

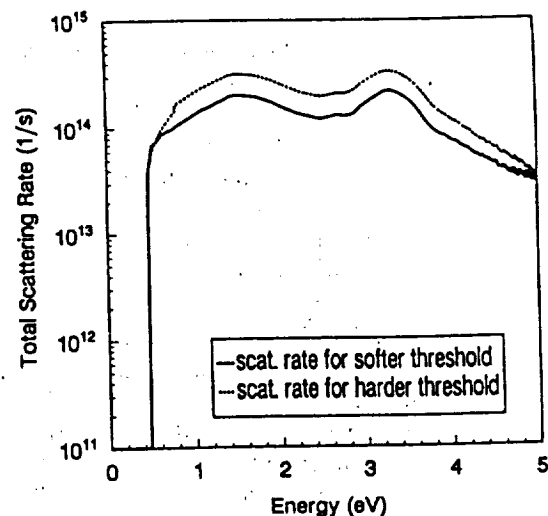


FIG. 3. Total calculated electron phonon scattering rate used in the soft threshold and hard threshold simulations. The high-energy scattering rate for the soft threshold case is matched at 0.5 eV with the phonon scattering rate calculated from Fermi's golden rule, while the rate for the hard threshold case is matched at 0.8 eV.

is determined by linearly interpolating between the values at each mesh point. If one or more of the mesh point values is zero, then the transition rate for the simulated electron is chosen equal to that of the nearest \mathbf{k} point. The impact ionization transition rate determined in this manner is then added to the total phonon scattering rate corresponding to the energy of the simulated electron. As in standard Monte Carlo algorithms, a random number is thrown to determine if the carrier is scattered. Given that the carrier scatters, a second random number is generated to determine the particular event that occurs out of all the possible events at that energy and \mathbf{k} vector.

The calculated interband impact ionization transition rate for the first two conduction bands of bulk GaAs as a function of k_x and k_y , holding k_z fixed at 0.0 within the reduced zone is presented in Figs. 1 and 2. As can be seen from comparison of these two figures, the transition rate is significantly greater in the second conduction band than in the first. The maximum transition rate in either band is only slightly greater than 10^{13} s^{-1} , significantly less than the competing phonon scattering rate at these energies. The probability that an ionization event will occur at any given time, relative to competing phonon scattering event, is then quite small. As a result, an electron can drift for some time to very high energy before suffering an impact ionization event.

III. CALCULATED RESULTS

At high carrier energies the phonon scattering rate is assumed, as stated above, to be dominated by deformation potential scattering. The scattering rate at these energies is determined by directly integrating over the final density of states calculated from the numerically generated band structure including collision broadening. A single deformation potential constant is selected in order to match the scattering

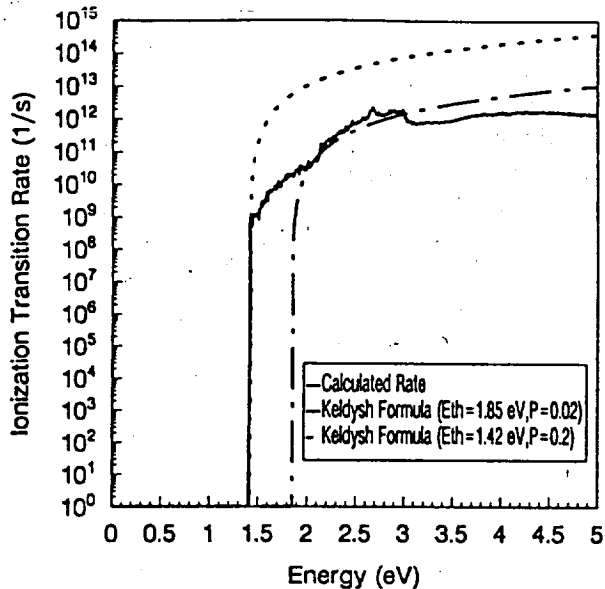


FIG. 4. Calculated impact ionization transition rates as a function of electron energy measured from the conduction-band minimum using the k -dependent and Keldysh formula with two different sets of P and E_{th} . The energy-dependent rate for the k -dependent rate is calculated by performing a further integration of the k -dependent rate.

rate calculated from Fermi's golden rule at a specific energy. Since the impact ionization coefficients are very sensitive to the total scattering rate, a small change in the total scattering rate will lead to a dramatic change in the impact ionization coefficients. Figure 3 shows the total electron phonon scattering rates with the deformation potential scattering rate matched to the scattering rate calculated from Fermi's golden rule at 0.5 and 0.8 eV, respectively. As is discussed later, the lower scattering rate, that matched at 0.5 eV, results in a softer threshold for the impact ionization process, while the higher scattering rate, that matched at 0.8 eV, results in a harder threshold. It should be noted that there is some arbi-

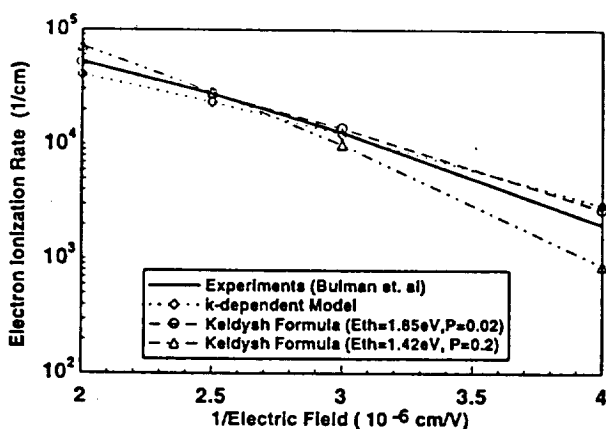
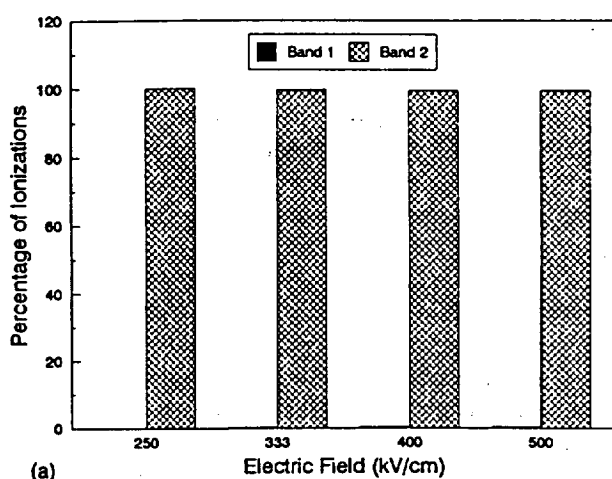
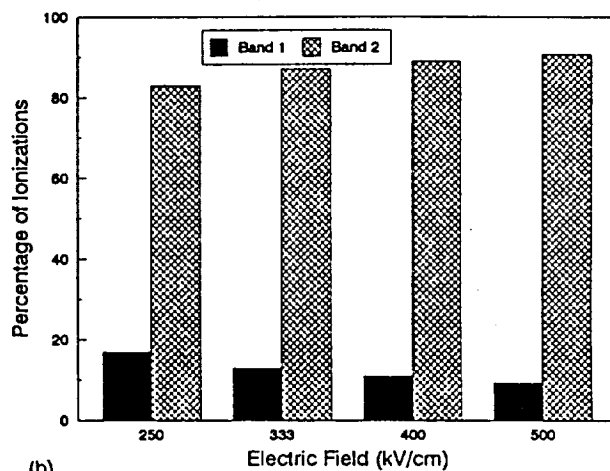


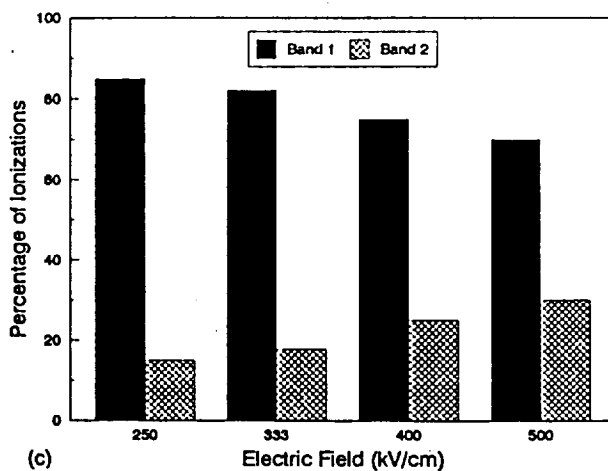
FIG. 5. Calculated and experimental results for the electron-impact ionization coefficients in bulk GaAs plotted as a function of inverse electric field, using the k -dependent model, the Keldysh formula model with $P=0.02$ and $E_{th}=1.85$ eV, and the Keldysh formula model with $P=0.2$ and $E_{th}=1.42$ eV. The experimental results are taken from Bulman *et al.* (Ref. 33).



(a)



(b)



(c)

FIG. 6. (a) Histogram showing the percentage of electron ionization events originating from the first and second conduction bands as a function of applied electric field based on the k -dependent model. (b) Histogram showing the percentage of electron ionization events originating from the first and second conduction bands as a function of applied electric field based on the Keldysh formula model with $P=0.02$, $E_{th}=1.85$ eV. (c) Histogram showing the percentage of electron ionization events originating from the first and second conduction bands as a function of applied electric field based on the Keldysh formula model with $P=0.2$, $E_{th}=1.42$ eV.

trariness in the total scattering rate since the deformation potentials are not fully known as a function of energy. Typically, the scattering rate is adjusted such that the calculated carrier drift velocities agree with the experimental measurements; however, the drift velocity is less sensitive to the scattering rate than the impact ionization coefficients. In these calculations, both sets of scattering rates show good agreement with the experimental saturation carrier drift velocity within the error tolerance of the Monte Carlo technique. Therefore, either total scattering rate is plausible.

An energy-dependent impact ionization transition rate can also be derived from the k -dependent rate by performing an additional integration over the initiating electron k -vector space, normalized by the density of states. The result is plotted as a function of initiating electron energy in Fig. 4, along with the rates using the Keldysh formula with two sets of P and E_{th} ; $P=0.02$ and $E_{th}=1.85$ eV, as well as $P=0.2$ and $E_{th}=1.42$ eV. Multiple sets of P and E_{th} may be found by fitting the calculated impact ionization coefficients to the experiment. The first set of values of $P=0.02$ and $E_{th}=1.85$ eV is chosen so as to match the Keldysh rate as close as possible to the energy-dependent rate derived from the k -dependent rate. The value of 0.02 for P is in the range of values which typically corresponds to a soft threshold condition. Another set of parameters for the Keldysh formula, $P=0.2$ and $E_{th}=1.42$ eV, is chosen to compare the difference between the hard and soft threshold models.

The impact ionization coefficients calculated using the k -dependent impact ionization transition rate, the Keldysh formula with $P=0.02$ and $E_{th}=1.85$ eV, and the Keldysh formula with $P=0.2$ and $E_{th}=1.42$ eV along with the experimental measurements of Bulman *et al.*³³ are plotted in Fig. 5 at an electric-field direction of [100]. The lower total scattering rate as shown in Fig. 3 is used in the calculations of both the k -dependent model and the Keldysh formula with $P=0.02$ and $E_{th}=1.85$ eV. In the case of the Keldysh formula with $P=0.2$ and $E_{th}=1.42$ eV the higher total scattering rate must be used due to the higher impact ionization transition rate and the harder threshold present in this case in order to agree with the experimental results. As can be seen from Fig. 5, the k -dependent calculations match the experiment fairly well throughout the range of applied electric-field strengths plotted here. This is excellent considering there is no adjustable parameter used in the model for the impact ionization transition rate. The only ambiguity in this case appears in the scattering rate. The calculated results using the Keldysh formula for the impact ionization transition rate, within the otherwise identical Monte Carlo model, are included for comparison. With two adjustable parameters, P and E_{th} it is possible to have better agreement with the experiments than the k -dependent model, as shown here for the case of $P=0.02$ and $E_{th}=1.85$ eV. The results calculated using the k -dependent model are closer to the results calculated by the Keldysh formula with a softer threshold as expected, since the k -dependent model also has effectively a very soft threshold.

The effect of the second conduction band on the impact ionization coefficients in bulk GaAs for all three impact ionization transition rate models has also been investigated. Fig-

ure 6(a) shows the percentage of electron ionization events originating from the first and second conduction bands within the k -dependent model. Nearly 100% of the impact ionization events originate from the second conduction band due to the fact that the impact ionization transition rate is much higher in the second conduction band. The percentage of electron ionization events originating from the first and second conduction bands for both Keldysh formula models are plotted in Figs. 6(b) and 6(c). In contrast to the k -dependent model, more ionization events originated from the first conduction band in both cases. This can be understood since in the Keldysh formula, the impact ionization transition rate is isotropic and no distinction is made between states in the first and second conduction bands. Therefore, the electrons have the same ionization probability at a given energy from the first conduction band as from the second conduction band in the Keldysh formula model, while in the k -dependent model the ionization probability is typically higher in the second conduction band than the first conduction band even at the same energy. Hence, it is not surprising that the Keldysh formula model predicts that more events will occur from the first conduction band than the k -dependent model does. As the threshold becomes lower and harder, more ionization events occur in the first conduction band. Therefore, the Keldysh formula model will predict very different physical results if a different set of P and E_{th} is used.

It is interesting to examine the impact ionization coefficients along different applied electric-field directions. The directional dependence of the impact ionization transition rate is incorporated naturally in the k -dependent formulation. Calculations of the electron-impact ionization coefficient at different applied electric field directions of [100], [110], and [111] show no field directional dependence of the impact ionization coefficients even with the k -dependent impact ionization transition rate. These results agree with previous calculations by Brennan and Hess³⁴ using the Keldysh formula and the calculations reported here using both Keldysh formula models. As shown in Fig. 6(a), almost all the impact ionization events originate from the second conduction band using the k -dependent formulation. Since the impact ionization transition rate in the second conduction band, shown in Fig. 2, is higher and more uniform than the rate in the first conduction band, little electric-field directional dependence of the electron-impact ionization coefficients is expected.

The electron number density functions, defined here as the product of the electron density of states function and the electron distribution function, are plotted in Figs. 7(a)–7(c) for all three models at an applied electric field of 250 kV/cm. The number density functions at the applied electric field of 500 kV/cm for all three models are presented in Figures 8(a)–8(c). Both the k -dependent model and the Keldysh formula model with the softer threshold predict similar electron distributions. The second peak in the distributions at an electric field of 500 kV/cm for the k -dependent model and the Keldysh formula with softer threshold indicate many electrons occupy the second conduction band at high applied electric fields. In the Keldysh formula model with the harder threshold, the distribution is much cooler. In this case, fewer

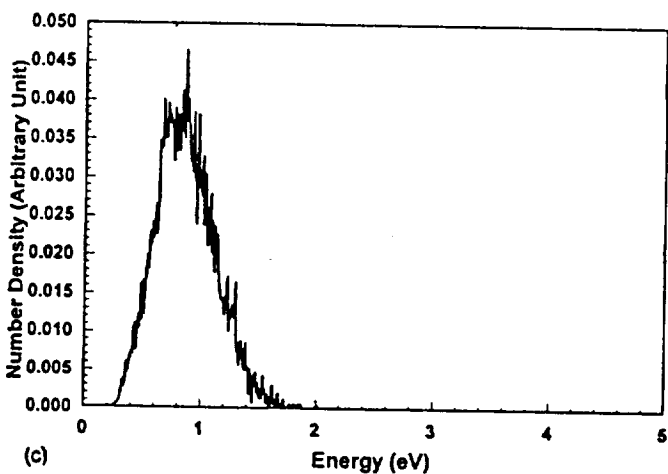
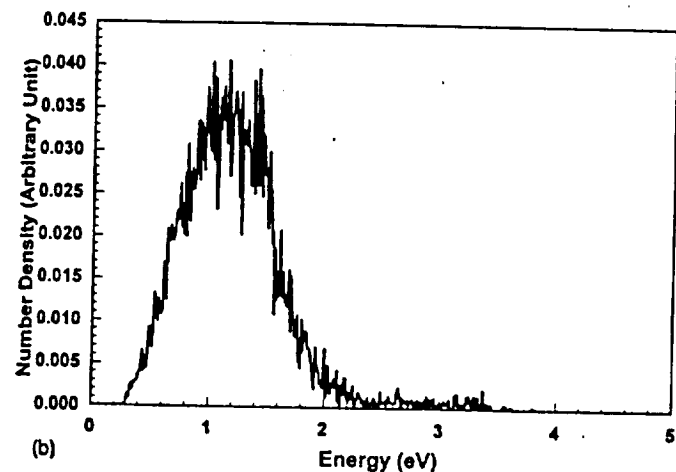
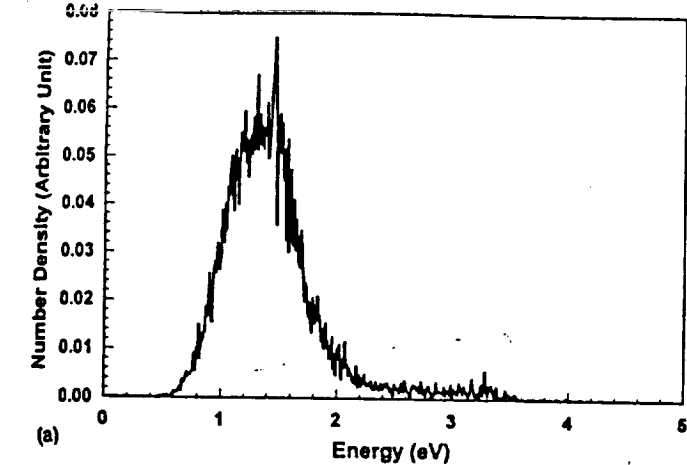


FIG. 7. (a) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 250 kV/cm calculated using the k-dependent model. Notice the significant number of electrons within the high-energy tail of the distribution. In the k-dependent model owing to the ultrasoft threshold present, the impact ionization process does not truncate the high-energy tail of the distribution. (b) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 250 kV/cm calculated using the Keldysh formula model with $P=0.02$, $E_{th}=1.85$ eV. As in the k-dependent model, the high-energy tail of the electron distribution is not truncated due to the soft threshold in the ionization process. (c) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 250 kV/cm calculated using the Keldysh formula model with $P=0.2$, $E_{th}=1.42$ eV. Notice that few electrons exist within the high-energy tail of the distribution in this case. The presence of a hard threshold acts to sharply truncate the high-energy tail at relatively low carrier energy.

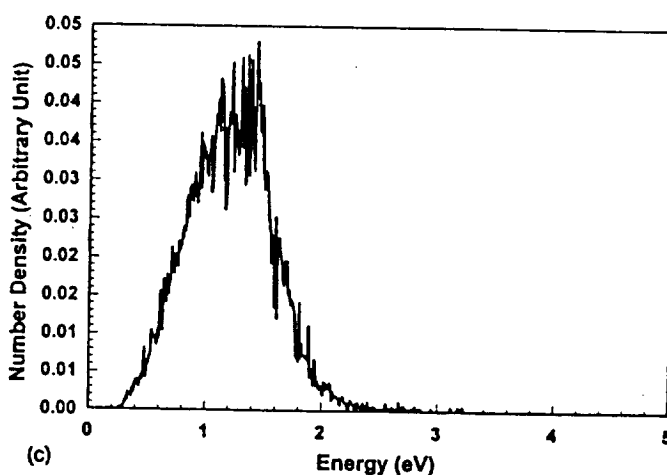
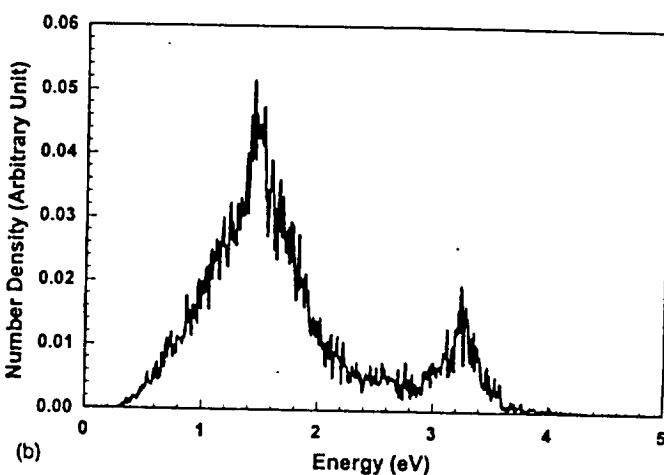
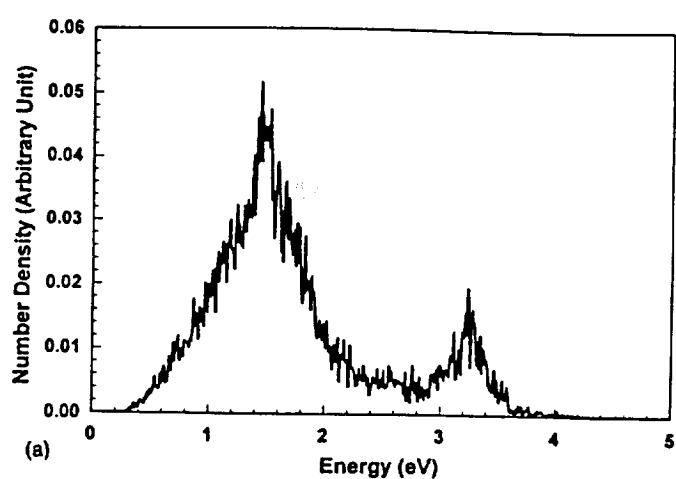


FIG. 8. (a) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 500 kV/cm calculated using the k-dependent model. The second peak in the distribution is due to carriers within the second conduction band. Notice that there is a significant population of electrons at very high energy due to the presence of a very soft threshold for impact ionization. (b) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 500 kV/cm calculated using the Keldysh formula model with $P=0.02$, $E_{th}=1.85$ eV. As in the k-dependent case, the second peak in the distribution is due to carriers within the second conduction band. Many electrons survive to high energy in this case owing to the very soft threshold for impact ionization. (c) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 500 kV/cm calculated using the Keldysh formula model with $P=0.2$, $E_{th}=1.42$ eV. Owing to the relatively hard threshold used in this case, the high-energy tail of the distribution is effectively truncated.

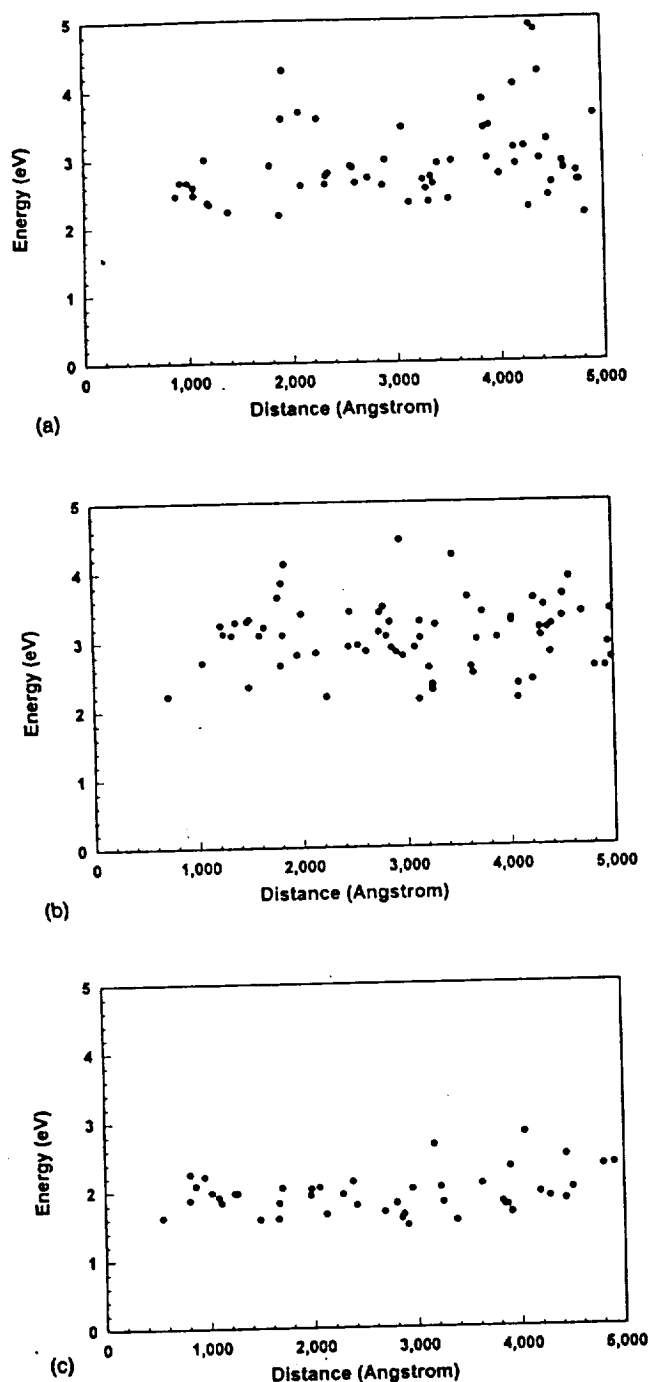


FIG. 9. (a) Scatter plot of electron energy and the location at which impact ionization occurs at the applied electric field of 333 kV/cm based on the k -dependent model. The dead space can be estimated by the distance measured from zero to the first ionization event. In this case, this distance is ~ 90.0 nm. Notice that the average energy at which an impact ionization event occurs is quite high, always greater than 2 eV and in many cases closer to 3–4 eV. (b) Scatter plot of electron energy and the location at which impact ionization occurs at the applied electric field of 333 kV/cm based on the Keldysh formula model with $P=0.02$, $E_b=1.85$ eV. As in the k -dependent case, the average energy at which an impact ionization event occurs is relatively high and the dead space is relatively long. (c) Scatter plot of electron energy and the location at which impact ionization occurs at the applied electric field of 333 kV/cm based on the Keldysh formula model with $P=0.2$, $E_b=1.42$ eV. Notice that the average energy at which an impact ionization event occurs is relatively low. Many ionizations occur at energies less than 2 eV and few carriers survive to energies much greater than 2 eV. Notice also that the dead space is predicted to be much less than in the k -dependent model. In this case, the dead space is ~ 50.0 nm almost half of that predicted for the k -dependent model.

electrons survive to high energies. This is, of course, due to the fact that in the hard threshold model, once a carrier attains threshold it is highly likely to suffer an impact ionization event. As a result, few carriers drift to energies significantly higher than the threshold.

The electron energies and positions at which impact ionizations occur are plotted in Figs. 9(a)–9(c) for all three models at the applied electric field of 333 kV/cm. Each dot in the figures represents an impact ionization event that occurred at energy E and distance D from launch. The electrons are all launched initially from zero energy. From inspection of the figures, the electron-impact ionization dead space can be estimated at each specific applied electric field. Notice that the dead space predicted by the harder threshold model is the shortest among the three models. This is because in the harder threshold model electrons can impact ionize at much lower energies [Fig. 9(c)] due to the lower threshold and the higher impact ionization probability once the carrier reaches the threshold. The distance traveled for the electrons to reach the ionization energy thus is much shorter. The k -dependent model has the longest dead space. In the k -dependent model most impact ionizations occur in the second conduction band, while in the Keldysh formula model electrons can more easily impact ionize within the first conduction band. Since electrons can only transfer to the second conduction band through scattering events in our simulation model, it takes a longer time and, thus, a longer distance, for the electrons to impact ionize. Due to the use of a higher and softer threshold, electrons need reach higher energy to impact ionize. This results in longer dead spaces than that reported in Ref. 35.

IV. CONCLUSIONS

In this article we have presented an analysis of the interband impact ionization rate in bulk GaAs using a semiclassical, k -dependent, numerical formulation of the impact ionization transition rate. It is found that the transition rate is substantially higher within the second conduction band than the first conduction band. As a result, the overwhelming majority of electrons impact ionize from states within the second conduction band than from the first conduction band, in contrast to what is predicted by a hard threshold, Keldysh formula model. In addition, the numerical, k -dependent formulation of the transition rate predicts that the electrons survive to very high energy, on average, before suffering an impact ionization event. Again, this is vastly different from that predicted by the hard threshold, Keldysh formula model wherein most of the electrons ionize at much lower energies.

Interestingly, both the hard threshold model, and the soft threshold models, i.e., either a soft threshold, Keldysh formula model or the numerical, k -dependent model, can be fashioned, by adjusting the total phonon scattering rate, to match the experimental results exceedingly well. Though the hard and soft threshold models require different total phonon scattering rates in their description, the two different rates examined here both lead to acceptable predictions of the steady-state electron drift velocities. As such, it is difficult under present circumstances to definitively choose which scattering rate is the most appropriate. A more detailed

analysis of the high-energy scattering rate, taking into account the energy dependence of the deformation potentials, needs ultimately to be done. Nevertheless, it is expected that the k -dependent transition rate formulation is far more accurate than the Keldysh formula. In order that the k -dependent transition rate calculations agree with the experimental measurements, the phonon scattering rate would have to be lower than that usually quoted for a hard threshold model.

If the threshold for interband impact ionization is indeed very soft in GaAs, as indicated by the present k -dependent transition rate calculations, the dead space for impact ionization is then significantly larger than previously estimated. A longer dead space may lead to different predictions of the electron-impact ionization enhancement in GaAs/AlGaAs multiquantum-well (MQW) structures. As a result, the optimal well and barrier widths of these devices may be quite different from those predicted using a harder threshold model. We are currently investigating how an ultrasoft threshold, as predicted by the k -dependent transition rate calculations presented here, would effect the predicted electron and hole ionization rates in MQW structures. These results will be presented elsewhere.

ACKNOWLEDGMENTS

This work was supported in part by ARPA and NASA through contract to NASA, Contract No. NAGW-2753. Additional support was received by the National Science Foundation through a Presidential Young Investigator Award made to K.B. and from Bell Northern Research through an industrial matching PYI award, Grant No. I25A3J. Computing support was provided by the Digital Equipment Corp. through Contract No. E21-H36.

¹W. Shockley, *Solid-State Electron.* **2**, 35 (1961).

²P. A. Wolff, *Phys. Rev.* **95**, 1415 (1954).

³H. Shichijo and K. Hess, *Phys. Rev. B* **23**, 4197 (1981).

⁴B. K. Ridley, *J. Phys. C Solid-State Phys.* **16**, 3373 (1983).

⁵J. Bude and K. Hess, *J. Appl. Phys.* **72**, 3554 (1992).

⁶R. Thoma, H. J. Peifer, W. L. Engl, W. Quade, R. Brunetti, and C. Jacoboni, *J. Appl. Phys.* **69**, 2300 (1991).

⁷J. S. Marsland, *Solid-State Electron.* **30**, 125 (1987).

⁸R. C. Woods, *Appl. Phys. Lett.* **52**, 65 (1988).

⁹J. Y. Tang and K. Hess, *J. Appl. Phys.* **54**, 5139 (1983).

¹⁰L. V. Keldysh, *Zh. Eksp. Teor. Fiz.* **48**, 1962 (1965) [*Sov. Phys. JETP* **21**, 1135 (1965)].

¹¹E. O. Kane, *Phys. Rev.* **159**, 624 (1967).

¹²N. Sano, T. Aoki, and A. Yoshii, *Appl. Phys. Lett.* **55**, 1418 (1989).

¹³N. Sano, M. Tomizawa, and A. Yoshii, *Appl. Phys. Lett.* **56**, 653 (1990).

¹⁴Y. Wang and K. F. Brennan, *J. Appl. Phys.* **75**, 313 (1994).

¹⁵J. Bude, K. Hess, and G. J. Iafrate, *Phys. Rev. B* **45**, 10 958 (1992).

¹⁶W. Quade, F. Rossi, and C. Jacoboni, *Semicond. Sci. Technol.* **7**, B502 (1992).

¹⁷P. D. Yoder, J. M. Higman, J. Bude, and K. Hess, *Semicond. Sci. Technol.* **7**, B357 (1992).

¹⁸E. Cartier, M. V. Fischetti, E. A. Eklund, and F. R. McFeely, *Appl. Phys. Lett.* **62**, 3339 (1993).

¹⁹R. Chin, N. Holonyak, Jr., G. E. Stillman, J. Y. Tang, and K. Hess, *Electron Lett.* **16**, 467 (1980).

²⁰F. Capasso, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Berr (Academic, New York, 1985), Vol. 22, Part D.

²¹F. Capasso, W. T. Tsang, A. L. Hutchinson, and G. F. Williams, *Appl. Phys. Lett.* **40**, 38 (1982).

²²F. Capasso, W. T. Tsang and G. F. Williams, *IEEE Trans. Electron Devices* **ED-30**, 381 (1983).

²³F. Y. Juang, U. Das, Y. Nashimoto, and P. K. Bhattacharya, *Appl. Phys. Lett.* **47**, 972 (1985).

²⁴H. Shichijo and K. Hess, *Phys. Rev. B* **23**, 4197 (1981).

²⁵Y. C. Chang, D. Z.-Y. Ting, J. Y. Tang, and K. Hess, *Appl. Phys. Lett.* **42**, 26 (1983).

²⁶M. V. Fischetti and S. E. Laux, *Phys. Rev. B* **38**, 9721 (1988).

²⁷K. F. Brennan, D. H. Park, K. Hess, and M. A. Littlejohn, *J. Appl. Phys.* **63**, 5004 (1988).

²⁸K. F. Brennan, N. Mansour, and Y. Wang, *Comput. Phys. Commun.* **67**, 73 (1991).

²⁹B. K. Ridley, *Quantum Processes in Semiconductors*, 2nd ed. (Oxford University Press, Oxford, 1988).

³⁰F. H. Pollak, C. W. Higginbotham, and M. Cardona, *J. Phys. Soc. Jpn.* **21**, 20 (1966).

³¹M. G. Burt, S. Brand, C. Smith, and R. A. Abram, *J. Phys. C* **17**, 6385 (1984).

³²C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer, Vienna, 1989).

³³G. E. Bulman, V. M. Robbins, K. F. Brennan, K. Hess, and G. E. Stillman, *IEEE Electron. Device Lett.* **EDL-4**, 181 (1983).

³⁴K. Brennan and K. Hess, *Solid-State Electron.* **27**, 347 (1984).

³⁵K. Brennan, *IEEE J. Quantum Electron.* **QE-24**, 2001 (1988).

Theoretical investigation of wave-vector-dependent analytical and numerical formulations of the interband impact-ionization transition rate for electrons in bulk silicon and GaAs

Ján Kolník, Yang Wang, Ismail H. Oğuzman, and Kevin F. Brennan
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0250

(Received 21 March 1994; accepted for publication 8 June 1994)

The electron interband impact-ionization rate for both silicon and gallium arsenide is calculated using an ensemble Monte Carlo simulation with the expressed purpose of comparing different formulations of the interband ionization transition rate. Specifically, three different treatments of the transition rate are examined: the traditional Keldysh formula, a new k -dependent analytical formulation first derived by W. Quade, E. Scholl, and M. Rudan [Solid State Electron. 36, 1493 (1993)], and a more exact, numerical method of Y. Wang and K. F. Brennan [J. Appl. Phys. 75, 313 (1994)]. Although the completely numerical formulation contains no adjustable parameters and as such provides a very reliable result, it is highly computationally intensive. Alternatively, the Keldysh formula, although inherently simple and computationally efficient, fails to include the k dependence as well as the details of the energy band structure. The k -dependent analytical formulation of Quade and co-workers overcomes the limitations of both of these models but at the expense of some new parameterization. It is found that the k -dependent analytical method of Quade and co-workers produces very similar results to those obtained with the completely numerical model for some quantities. Specifically, both models predict that the effective threshold for impact ionization in GaAs and silicon is quite soft, that the majority of ionization events originate from the second conduction band in both materials, and that the transition rate is k dependent. Therefore, it is concluded that the k -dependent analytical model can qualitatively reproduce results similar to those obtained with the numerical model yet with far greater computational efficiency. Nevertheless, there exist some important drawbacks to the k -dependent analytical model of Quade and co-workers: These are that it does not accurately reproduce the quantum yield data for bulk silicon, it requires determination of a new parameter, related physically to the overlap integrals of the Bloch state which can only be adjusted by comparison to experiment, and fails to account for any wave-vector dependence of the overlap integrals. As such the transition rate may be overestimated at those points for which "near vertical," small change in k , transitions occur.

I. INTRODUCTION

The theoretical study of interband impact ionization has been greatly aided by the advent of numerical methods. Owing to the complexity of the energy band structure at high energies, at which impact ionization is initiated, as well as the complicated carrier-phonon scattering mechanisms, simplified analytical formulations, such as the early theories of Shockley,¹ Wolff,² and Baraff³ have limited validity. Numerical models for calculating the ionization rate were then advanced based on the Monte Carlo method;^{4,5} however, these approaches utilized parabolic or nonparabolic analytical energy bands which are of questionable validity at high carrier energies. The full details of the energy band structure were first accounted for in the Monte Carlo model of Shichijo and Hess.⁶ In their model the dynamics of the electrons in bulk GaAs were simulated within the first conduction band, calculated based on an empirical pseudopotential model. Later, Fischetti and Laux⁷ developed a more advanced Monte Carlo simulator for studying impact ionization which improved the numerical accuracy as well as incorporated transport within higher conduction bands.

In order to calculate the overall impact ionization rate using the Monte Carlo technique, it is necessary to formulate

an expression for the impact ionization transition rate. This formulation is then incorporated into the Monte Carlo simulator and is treated as an additional scattering mechanism. An impact-ionization event is chosen stochastically in the usual way⁸ through the use of a random number based on the relative magnitude of the ionization transition rate compared to competing phonon scattering events. Although there has been recent work on formulating the ionization transition rate using a higher-order quantum-mechanical perturbation theory expansion,⁹⁻¹¹ the impact-ionization transition rate is more typically determined from use of Fermi's golden rule. Fermi's golden rule can be expressed as¹²

$$W_{ii} = \int \frac{2\pi}{\hbar} |M|^2 \delta(E_f - E_i) \delta(\Delta k) dS_f, \quad (1)$$

where dS_f represents integration over all the final states, E_i and E_f are the energies of the initial and final states, respectively, Δk is the momentum change during the interaction. The matrix element M in Eq. (1) is formed using both the direct M_d and exchange M_e terms following Ridley¹² as

$$|M|^2 = |M_d|^2 + |M_e|^2 + |M_d - M_e|^2, \quad (2)$$

where

$$M_d = \frac{e^2}{\epsilon V} \frac{I(\mathbf{k}_1, \mathbf{k}'_1)I(\mathbf{k}_2, \mathbf{k}'_2)}{|\mathbf{k}'_1 - \mathbf{k}_1|^2 + \lambda^2},$$

$$M_e = \frac{e^2}{\epsilon V} \frac{I(\mathbf{k}_1, \mathbf{k}'_2)I(\mathbf{k}_2, \mathbf{k}'_1)}{|\mathbf{k}'_2 - \mathbf{k}_1|^2 + \lambda^2}. \quad (3)$$

In Eq. (3), $\mathbf{k}_1, \mathbf{k}'_1$ are the wave vectors of the incident electron before and after the interaction, $\mathbf{k}_1, \mathbf{k}'_2$ represent the states of the secondary electron-hole pair after the collision, λ is the static screening factor, and e is the electronic charge; V is the crystal volume, and ϵ is the dielectric function. Quantities $I(\mathbf{k}_1, \mathbf{k}'_1), I(\mathbf{k}_2, \mathbf{k}'_2), I(\mathbf{k}_1, \mathbf{k}'_2), I(\mathbf{k}_2, \mathbf{k}'_1)$ are overlap integrals.¹² Following Ridley,¹² the term $|M_d - M_e|^2$ is neglected and the direct and the exchange terms are assumed to be equal.

In most previous Monte Carlo simulations (e.g., Refs. 6 and 13) the impact-ionization transition rate is calculated using the Keldysh formula,¹⁴ which can be obtained from Eq. (1) by adopting several simplifying assumptions: These are that the transition is calculated assuming a direct semiconductor with parabolic bands for all of the carriers, the overlap integrals are considered to be constant and the denominator in the matrix element in Eq. (3) is also taken as a constant with thresholds values of $\mathbf{k}_1, \mathbf{k}'_1$. This yields the very well-known quadratic dependence of the transition rate W_{ii} on the energy E of the initiating particle, usually expressed as

$$W_{ii} = pW(E_{th}) \left(\frac{E - E_{th}}{E_{th}} \right)^2, \quad (4)$$

where E_{th} is the threshold energy and $W(E_{th})$ is the total phonon scattering rate at threshold. E_{th} and the prefactor p cannot generally be determined from first principles; instead, they are determined through comparison of the calculations to experimental data. Aside from this obvious limitation of the Keldysh formula, the neglect of an accurate accounting of the high-energy region of the band structure also results in the failure to properly assess the \mathbf{k} dependence of the ionization transition rate itself. Recent theoretical studies have indicated that there is a significant deviation in the ionization transition rate with respect to the initiating carrier's \mathbf{k} vector in different materials.^{15,16}

Several alternative approaches to the Keldysh formula for calculating the interband impact ionization rate have been recently presented.¹⁵⁻²¹ In each of these methods, the interband impact-ionization transition rate is evaluated numerically and can be then incorporated into an ensemble Monte Carlo calculation. Although these approaches presently offer a far more accurate treatment of impact ionization than the Keldysh formula, they require extensive numerical computation. In the approach adopted by Wang and Brennan^{15,16} the transition rate is determined directly from Fermi's golden rule by numerically integrating Eq. (1) at different \mathbf{k} points within the first Brillouin zone. Their analysis includes the direct calculation of the overlap integrals from the numerically generated wave functions using a $\mathbf{k} \cdot \mathbf{p}$ method. The transition rate is evaluated by integrating over several million final states directly evaluating the overlap integrals in each case. This is particularly computationally intensive since the overlap integrals must be evaluated for each final

state. Sano and Yoshii¹⁷ have used a less computationally intensive scheme. In their method, the three-dimensional \mathbf{k} space incorporating the reduced zone is partitioned into small cubes. The matrix elements within each cube are assumed to be equal alleviating the need to evaluate the overlap integrals for each possible state. In this way, the number of computations can be greatly reduced at the expense of increased memory requirements. However, we have found that the overlap integrals vary in a nonpredictive manner from point to point in our $\mathbf{k} \cdot \mathbf{p}$ calculations and cannot generally be assumed to be the same even for states very close in \mathbf{k} space. For this reason and memory storage limitations, in the model of Wang and Brennan^{15,16} the overlap integrals are evaluated for every final state. Nevertheless, all of the numerical approaches indicate that the transition rate is dependent on the initiating electron wave vector.

Recently, another analytical treatment of expression (1), which also includes the \mathbf{k} -vector dependence of the impact-ionization transition rate, was presented by Quade, Scholl, and Rudan.²² They calculated the impact-ionization transition rate for the general case of an arbitrarily shaped energy band for the impact-ionizing conduction electron, and three anisotropic parabolic bands for the final states with their extrema located at different points in the \mathbf{k} -vector space. The assumption of the parabolic energy bands for the final states is acceptable since the final states are generally at low energy near the band minimum where the parabolic approximation is typically good. Other approximations used in the derivation include nondegeneracy and a constant value for the overlap integrals. Although the formulation developed by Quade and co-workers²² contains some parameterization due to the fact that the overlap integrals are not evaluated but are assumed to be constant parameters, it offers a far more computationally efficient means of evaluating the ionization transition rate with the inclusion of the \mathbf{k} dependence than the numerical method of Wang and Brennan.^{15,16} Subsequently, it is of interest to examine how the \mathbf{k} -dependent, analytical model of Quade and co-workers²² compares to the numerical model of Wang and Brennan.^{15,16}

In this article, we incorporate the formulation of the impact-ionization transition rate of Quade and co-workers²² within an ensemble Monte Carlo simulator to determine the electron impact-ionization rate in bulk silicon and GaAs. The impact-ionization rate calculated in this way is then compared to experimental data as well as calculations made in an otherwise identical Monte Carlo simulator using the Keldysh formula and the direct, \mathbf{k} -dependent transition rate. The nature of the ionization rate is further probed using the Quade and co-workers' model to determine if it yields similar physical results to that of the direct, numerical model. Specifically, the transition rate within each band is calculated and the percentage of ionization events originating from each conduction band is determined. The number density distributions and the quantum yield as a function of energy are determined and compared between the different theoretical models as well. In Sec. II the details of the theoretical models are reviewed. The calculated results are presented in Sec. III and conclusions are drawn in Sec. IV.

II. MODEL DESCRIPTION

The total impact-ionization rate as a function of inverse electric field is calculated using an ensemble Monte Carlo simulation, based on the original work of Shichijo and Hess.⁶ The details of the simulator have been extensively described elsewhere;^{15,16} therefore, only the main features are outlined here. The simulation includes the full details of the first and second conduction bands for silicon or gallium arsenide. The scattering rate in either case is determined using the improved phonon scattering technique of Chang *et al.*²³ and the improved band-structure interpolation technique of Fischetti and Laux.⁷ The total phonon scattering rate is determined in the following way. Within the low-energy region, the scattering rate is calculated from Fermi's golden rule for all the relevant mechanisms present.²⁴ Within the high-energy range, deformation potential scattering is assumed to be the dominant scattering mechanism. The scattering rate in this region is obtained by integrating over the final density of states, calculated from the numerically generated band structures, including collision broadening of the final state. The deformation potential is assumed constant and is selected to match the scattering rate calculated from Fermi's golden rule at a low specific energy. Interband electronic transitions are enabled via the action of the deformation potential scattering as described in Ref. 15. The actual scattering rates used in these calculations have been published previously in Ref. 15 for silicon and Ref. 16 for GaAs.

The impact-ionization transition rate is formulated in three different ways and is then incorporated into the Monte Carlo simulator described above. The different transition rate formulations used are a direct, numerical method previously described in detail in Refs. 15 and 16, the Keldysh formula described in detail in Refs. 6 and 7, and a new k -dependent analytical formulation derived by Quade and co-workers²² which is summarized below.

The k -dependent analytical formulation of the impact-ionization transition rate, hereafter referred to as the Quade formula, predicts the ionization transition rate for an initiating electron of wave vector k_1 , colliding with another electron in the valence band, to be²²

$$W_{ii}(k_1)_{\alpha,\beta,\gamma} = \int_{-\infty}^{K_1^2} d\xi W_{ii}(K_1, \xi)_{\alpha,\beta,\gamma} \delta[g(k_1) - \xi], \quad (5)$$

where

$$W_{ii}(K_1, \xi)_{\alpha,\beta,\gamma} = \frac{1}{\tau} \left[\frac{1}{2} \left(\frac{K_1}{\sqrt{\xi_\lambda}} + \frac{\sqrt{\xi_\lambda}}{K_1} \right) - 1 \right] \Theta(K_1^2 - \xi), \quad (6)$$

$$\frac{1}{\tau} = \frac{1}{4\pi^2} \frac{\sqrt{\mu}}{\prod_{i=1}^3 (\alpha_i + \beta_i)^{1/2}} \left(\frac{e^2}{\epsilon} \right)^2 \frac{m_\delta F}{\hbar^3}, \quad (7)$$

$$\mu_i = \frac{\alpha_i \beta_i + \alpha_i \gamma_i + \beta_i \gamma_i}{\alpha_i + \beta_i}, \quad (8)$$

$$g(k_1) = \frac{1}{\mu} \left(E_0 + \sum_{i=1}^3 \frac{\alpha_i \beta_i}{\alpha_i + \beta_i} (Z_{\alpha,i} - Z_{\beta,i})^2 + \sum_{i=1}^3 \gamma_i (k_{1,i}^L - Z_{\gamma,i})^2 - E(k_1^L) \right), \quad (9)$$

$$K_{1,i} = \frac{1}{\sqrt{\mu \mu_i}} \left(\gamma_i (k_{1,i}^L - Z_{\gamma,i}) - \frac{\alpha_i \beta_i}{\alpha_i + \beta_i} (Z_{\alpha,i} - Z_{\beta,i}) \right), \quad (10)$$

and

$$\xi_\lambda = \frac{1}{2} [\sqrt{(\lambda^2 - \xi)^2 + 4K_1^2 \lambda^2} - (\lambda^2 - \xi)]. \quad (11)$$

In the above set of equations, α , β , and γ represent a parabolic valence band and two possibly different parabolic conduction bands for the final states of the hole and electrons after the impact ionization event, respectively. The anisotropic bands are parametrized by ratios of effective masses $\alpha_i = m_\delta / m_{\alpha,i}$, $\beta_i = m_\delta / m_{\beta,i}$, $\gamma_i = m_\delta / m_{\gamma,i}$, where i labels the Cartesian coordinate, and the extremal points of each band are positioned at Z_α , Z_β , Z_γ ; m_δ is an isotropic effective mass of the conduction band for the initiating electron. F represents the product of the squared overlap integrals and is taken as constant for each initiating band but may vary between bands. μ stands for the geometrical average of the quantities μ_1 , μ_2 , μ_3 defined by Eq. (8). The arbitrarily shaped function $E(k_1)$ represents the energy of the initiating electron, E_0 represents the gap plus offsets of the extremal points of the bands α , β , γ . $k_1^L = k_1 - L$, where L is a reciprocal lattice vector and Θ is the Heaviside function. Equation (5) is derived assuming a static dielectric constant. Although recent work¹⁷ has shown that a full wave-vector, frequency dependent dielectric function can modify the ionization transition rate, its incorporation in the Quade formula would not lead to an analytical expression. For this reason, a constant dielectric function is used within the Quade formulation. To facilitate comparison of the Quade model to other models, a constant dielectric function is used within the numerical and Keldysh formula models as well.

The total impact-ionization scattering rate for the initiating particle with wave vector k_1 is calculated taking into account all of the possible final states within the full Brillouin zone. The total impact-ionization transition rate can then be expressed as

$$W_{ii}(k_1) = \sum_{\alpha} \sum_{\beta} \sum_{\gamma} W_{ii}(k_1)_{\alpha,\beta,\gamma}, \quad (12)$$

where the summation goes over all the valence bands α and all possible valleys β , γ in the first and second conduction bands. The summations in Eq. (12) are taken so as to include both direct and exchange terms.

The material parameters used in the calculations have been taken from Refs. 25–28 for silicon and Refs. 29–32 for gallium arsenide. For the case of the second conduction band in both materials, the longitudinal m_l and transverse m_t effective masses for the X valleys have been obtained from the numerically calculated band structures determined by a k -calculation. The longitudinal and transverse masses used for silicon and gallium arsenide in the second conduction band

TABLE I. Energies of symmetry points (relative to the top of the valence band, in eV) and values of relative effective masses used in the calculations of the impact-ionization transition rate in silicon and gallium arsenide.

	Silicon	Gallium arsenide
$E(\Gamma_{CI})$	3.42 ^a	1.423 ^b
$E(X_{CI})$	1.12 ^a	1.898 ^b
$E(L_{CI})$	1.92 ^a	1.706 ^b
$E(X_{CII})$	1.14 ^c	1.973 ^c
$E(\Gamma_{SO})$	-0.044 ^d	-0.340 ^b
$m_r(X_{CI})$	0.19 ^e	0.27 ^f
$m_r(X_{CI})$	0.92 ^e	1.98 ^f
$m_r(X_{CII})$	0.21 ^e	0.23 ^c
$m_r(X_{CII})$	0.29 ^e	0.40 ^e
$m(L_{CI})$	0.284 ^e	0.18 ^f
$m(\Gamma_{CI})$...	0.067 ^f
$m(\Gamma_{HH})$	0.537 ^d	0.45 ^f
$m(\Gamma_{LH})$	0.153 ^d	0.082 ^f
$m(\Gamma_{SO})$	0.234 ^d	0.170 ^f

^aReference 32.

^bReference 30.

^cValues obtained from the numerically calculated band structure.

^dReference 26.

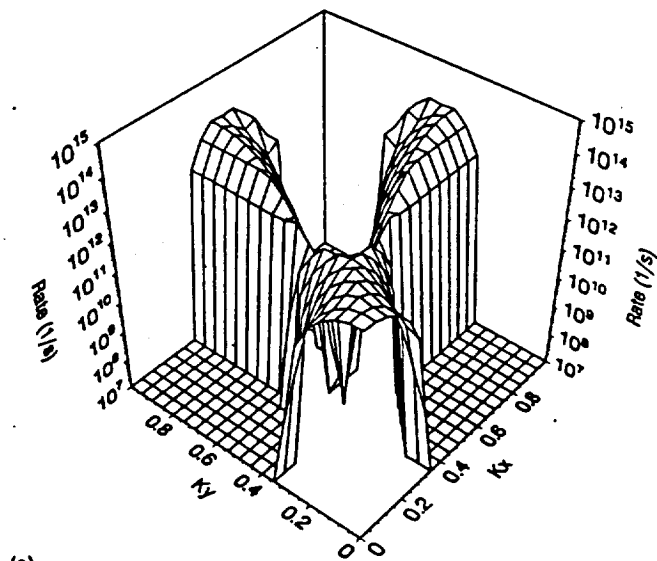
^eReference 25.

^fReference 31.

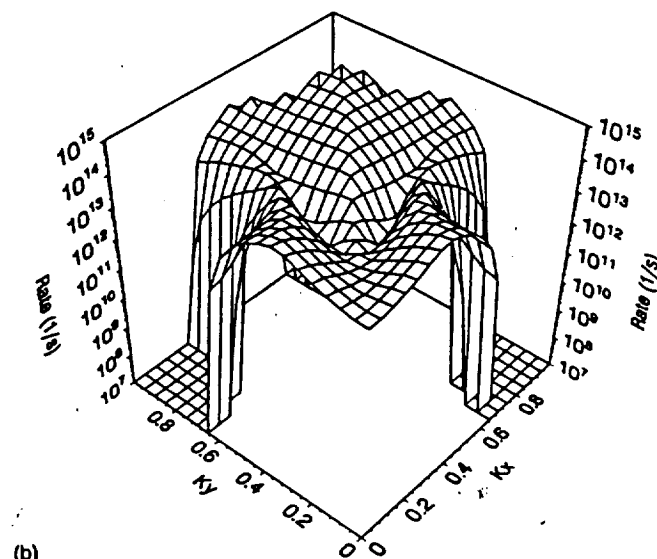
are reported in Table I. In the case of L valleys, Γ valleys, and all the valence bands, the isotropic effective masses have been used and their values are also listed in Table I, as well as the energies of symmetry points relative to the top of the valence band.

The interband impact-ionization transition rate calculated using the approach based on Quade and co-workers' results assuming that the overlap integrals are simply unity is presented in Figs. 1(a) and 1(b) for the first and second conduction bands in silicon, and in Figs. 2(a) and 2(b) for the first and second conduction bands in gallium arsenide, respectively. Comparison of Figs. 1 and 2 to similar figures obtained by Wang and Brennan^{15,16} using the completely numerical technique in which the overlap integrals are evaluated directly shows good qualitative agreement between the two approaches. For bulk silicon the primary difference, aside from the overall magnitude, is that the number of points for which a nonzero transition rate has been calculated is higher using the completely numerical approach. A similar effect can be seen in the case of the first and second conduction bands in gallium arsenide. Again notice that there is a significant difference in the number of nonzero transition rate points between the two models (data calculated using the numerical approach can be found in Ref. 16) for the first conduction band; however, the rate from these points is relatively small compared to the competing scattering rates.

Further examination of Figs. 1 and 2 reveals that the transition is greater for electrons originating from the second conduction band than from the first conduction band, which corresponds to the results based on the numerical calculations.^{15,16} This is particularly true for GaAs. As is discussed below, due to the fact that the transition rate is greatest for the second conduction band, most ionization events occur from electrons which originate from within the second conduction band.



(a)

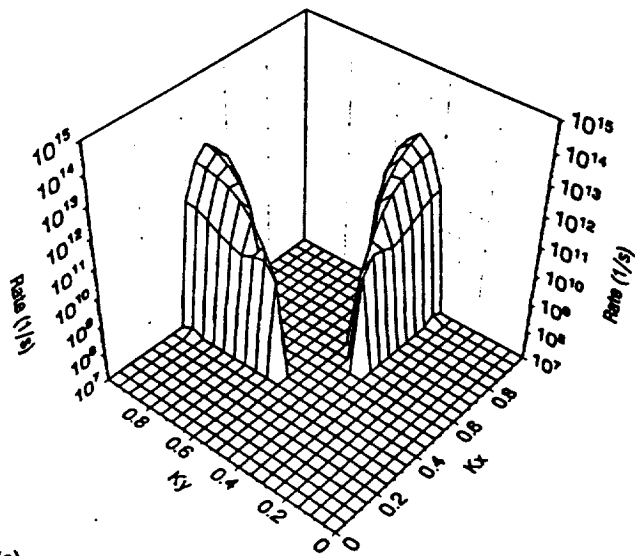


(b)

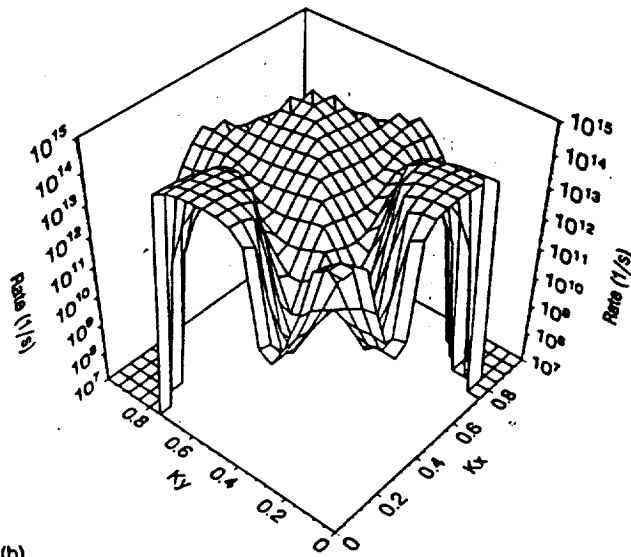
FIG. 1: Calculated interband impact-ionization transition rate for the first and second conduction band in bulk silicon as a function of k_x and k_y (k_z is fixed at 0.0) within the reduced zone of the first Brillouin zone obtained by the analytical formulation based on the Quade model. Initiating electron is in (a) the first conduction band and (b) the second conduction band.

III. RESULTS

The impact-ionization rate as a function of inverse electric field, calculated using the ensemble Monte Carlo simulation, is presented in Figs. 3 and 4 for silicon and gallium arsenide, respectively. In each case, the rate is calculated using the Keldysh formula, the k -dependent numerical formulation, and the Quade formula for the impact-ionization transition rate. The experimentally obtained data are also shown in both figures. The experimental results of van Overstraeten and DeMan,³³ Grant,³⁴ and Woods, Johnson, and Lambert³⁵ for electron-initiated impact ionization in silicon are plotted in Fig. 3. These data comprise a representative set of data which cover the full range of experimental measurements for electron-impact ionization in silicon. The experimental measurements of Bulman *et al.*³⁶ are plotted for gallium arsenide in Fig. 4.



(a)



(b)

FIG. 2. Calculated interband impact-ionization transition rate for the first and second conduction band in bulk GaAs as a function of k_x and k_y (k_z is fixed at 0.0) within the reduced zone of the first Brillouin zone obtained by the analytical formulation based on the Quade model. Initiating electron is in (a) the first conduction band and (b) the second conduction band.

In the case of silicon, good agreement with the data of Woods and co-workers is obtained for the k -dependent numerical formulation, as was also reported previously.¹⁵ No adjustable parameters for the impact-ionization transition rate are used in this case. Very similar results can be obtained using the Quade formulation by choosing the values for the parameter F in Eq. (7) equal to 0.24 for both the first and second conduction bands. F corresponds to an "average" value of the squared overlap integrals. The squared overlap integrals always have value between 0 and 1. In the analytical model of Quade and co-workers²² and that presented here, F is treated as a parameter which is chosen so as to yield good agreement with experiment. Best agreement with the low range of silicon experimental data occurs for $F=0.24$ given the phonon scattering rate adopted here. The calculated silicon impact-ionization rate obtained using the

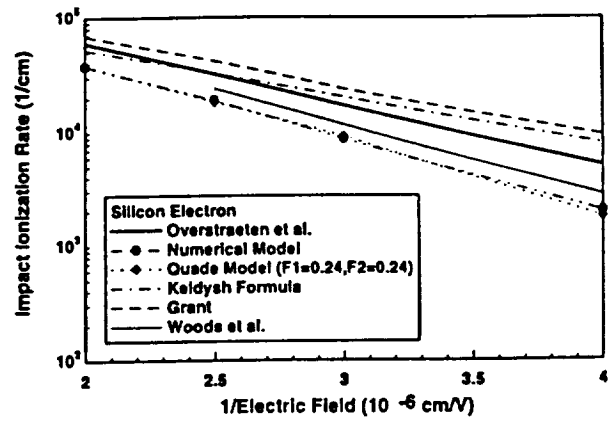


FIG. 3. Calculated and experimental results for the electron impact-ionization coefficients in bulk silicon plotted as a function of inverse electric field. Results of the calculations of the ionization rate including the Keldysh formula, k -dependent numerical, and Quade model are presented. Experimental results of Refs. 33, 34, and 35 are also shown. F_1 and F_2 represent the values of the averaged squared overlap integrals used in the Quade model for bands 1 and 2, respectively.

Keldysh formula with $p=0.02$ and $E_{th}=1.13$ eV is also shown in Fig. 3. A value of p of about 0.02 is typically considered to correspond to a soft threshold. In contrast to both k -dependent models, better agreement with the higher range of experimental measurements can be seen for this particular choice of threshold and p ; however, a different choice of the parameter F within the Quade formulation would also provide a better fit to the higher range of experimental data as well.

In gallium arsenide the dependence of the impact-ionization rate on the inverse electric field obtained using the numerical, the Quade models with $F=0.05$ and 0.005 for the first and second conduction band, respectively, and the Keldysh formula with $p=0.02$, $E_{th}=1.85$ eV for the impact-ionization transition rate, shows in all three cases fairly good agreement with the experimentally measured data

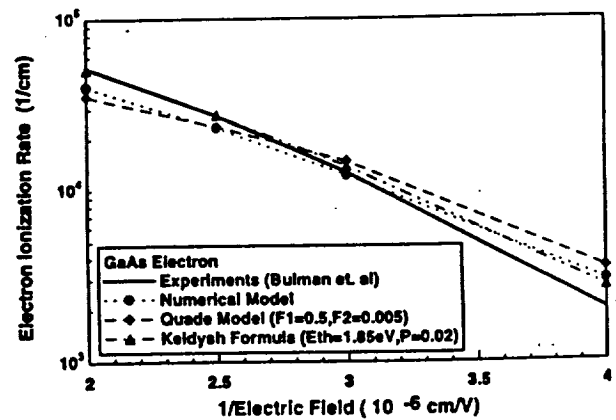


FIG. 4. Calculated and experimental results for the electron impact ionization coefficients in bulk gallium arsenide plotted as a function of inverse electric field, using the Keldysh formula, k -dependent numerical and Quade model. Experimental results shown are those of Bulman *et al* (Ref. 36). F_1 and F_2 represent the values of the averaged squared overlap integrals used in the Quade model for bands 1 and 2, respectively.

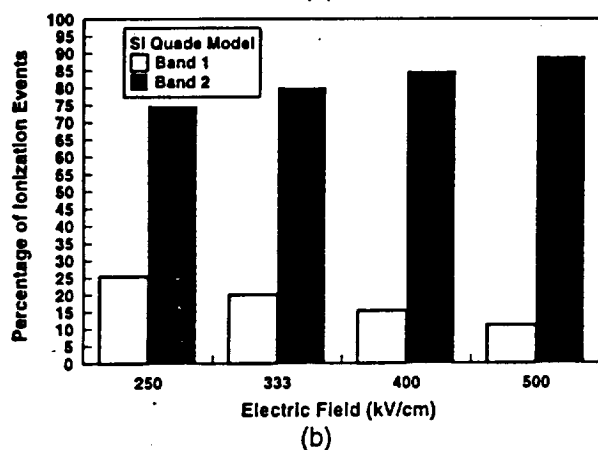
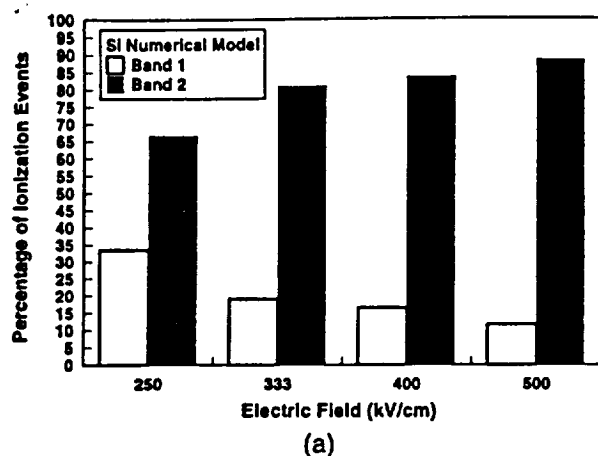


FIG. 5. Histogram showing the percentage of ionization events originating from the first and second conduction bands in silicon for (a) the k -dependent numerical formulation and (b) the Quade formulation.

(Fig. 4). It should be noted that in the case of the approach based on the Quade formula, the selection of the fitting factor F for the first conduction band essentially does not play any role due to the negligible impact-ionization transition rate of the electrons in the first conduction band with respect to the second conduction band.

The percentage of ionization events originating from the first and the second conduction band in bulk silicon as a function of applied electric field is plotted in Fig. 5(a) for the numerical model and in Fig. 5(b) for the model based on the Quade formula. At the applied electric fields investigated here the impact-ionization events originate predominantly in the second band in both cases. Therefore, on the basis of these reports, it appears that most electron-initiated ionization events originate from electrons within the second conduction band. The number of events originating in the first conduction band is found to be higher for the numerical model than for the model using the Quade formula at the field of 250 kV/cm, while for higher fields this difference vanishes. This slightly different behavior of the two models at low fields is caused by the fact that the threshold energy for impact ionization is about 1.32 eV in the Quade model for the particle originating in the first conduction band and 2.11 eV for the second conduction band, while for the numerical model the threshold energy value is not well defined.

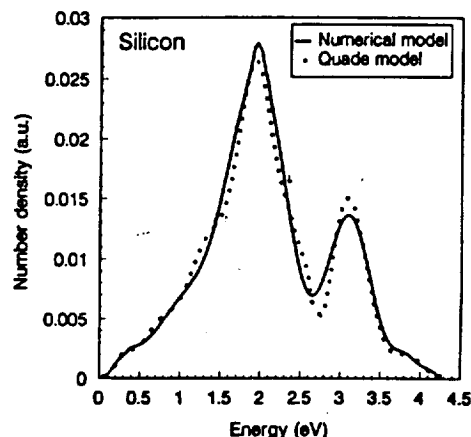


FIG. 6. The electron number density function in silicon, defined as the product of the density-of-states function and the electron distribution function, at an applied electric field of 500 kV/cm, calculated using the k -dependent numerical model and the Quade model.

Instead, in the numerical model ionization is possible at all states with energy greater than the band gap which are found to satisfy the momentum and energy conservation conditions. Therefore, there is no sharp threshold energy, other than the band gap, below which ionization cannot occur and above which ionization does occur. Thus, the contribution of the first conduction band at low fields can be expected to be higher when the numerical model is used.

The effect of the second conduction band on the impact-ionization rate in gallium arsenide is even more important than in silicon. The impact-ionization transition rate for electrons in the first conduction band is much lower than the rate within the second conduction band in both models as can be seen from Figs. 2(a) and 2(b) and from Ref. 16. As a result, virtually all the electrons which impact ionize originate within the second conduction band. The effect is even stronger using the Quade formulation than the numerical formulation. Though the value of F used in the analytical formulation is unknown *a priori*, i.e., without first comparing the calculations to experiment, it should be noted that F can never be greater than 1. Therefore, given that the transition rate of electrons within the first conduction band is several orders of magnitude lower than from the second conduction band with $F = 1$, as shown by Figs. 2(a) and 2(b), it is clear that the first conduction band plays no significant role in impact ionization in bulk GaAs for reasonable values of F . Hence, in either the numerical or analytical k -dependent models the predominant source of ionizing electrons in bulk GaAs is carriers within the second conduction band.

The electron number density function defined as the density-of-states function multiplied by the electron distribution function, is shown in Fig. 6 for silicon at an applied electric field of 500 kV/cm for both the numerical k -dependent model and the Quade model. In both cases, the number of electrons which survive to high energies is relatively high. The second peak in each of the figures corresponds to the electron population in the second conduction band and is slightly larger for the Quade model. Both models essentially confirm previous results³⁷⁻³⁹ that the threshold in

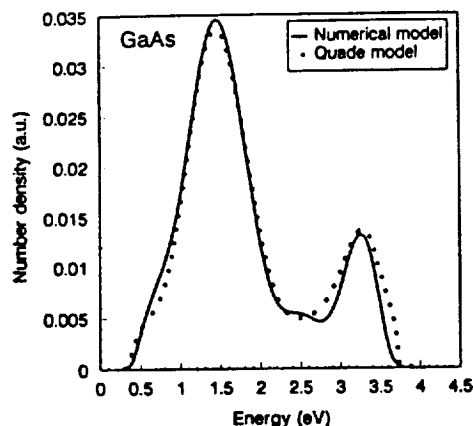


FIG. 7. The electron number density function in gallium arsenide, defined as the product of the density-of-states function and the electron distribution function, at an applied electric field of 500 kV/cm, calculated using the k-dependent numerical model and the Quade model.

silicon is exceedingly soft and that the electrons can reach energies substantially higher than the band gap before they impact ionize.

The electron number density functions in gallium arsenide are plotted in Fig. 7 calculated using the k-dependent numerical and Quade models for the applied electric field of 500 kV/cm. In both cases the distribution functions are similar and the effect of the second conduction band (the second peak at higher energies) is again slightly larger in the Quade model. Notice that the number distribution remains substantially large at high carrier energies indicating an exceedingly soft effective threshold for impact ionization in GaAs. Again these results are consistent with previously reported observations.²⁹

The energy-dependent impact-ionization transition rate, derived from the k-dependent rate in a manner described in Ref. 15, is plotted in Fig. 8 for the numerical k-dependent formulation and the Quade formula, respectively, for silicon.

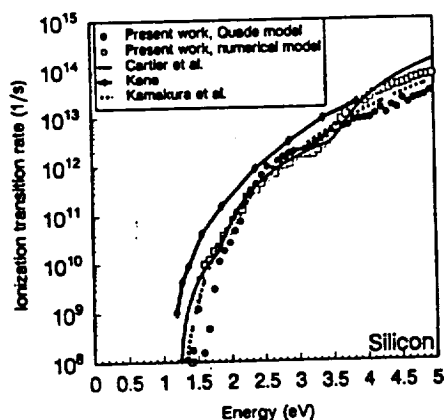


FIG. 8. Calculated impact-ionization transition rate in silicon as a function of electron energy measured from the conduction-band minimum. The energy-dependent rate is calculated by performing a further integration of the k-dependent rate following the approach explained in Ref. 15 for the numerical and Quade models; the results of Kane (Ref. 40) and Cartier *et al.* (Ref. 41) and Kamakura *et al.* (Ref. 19) are also shown.

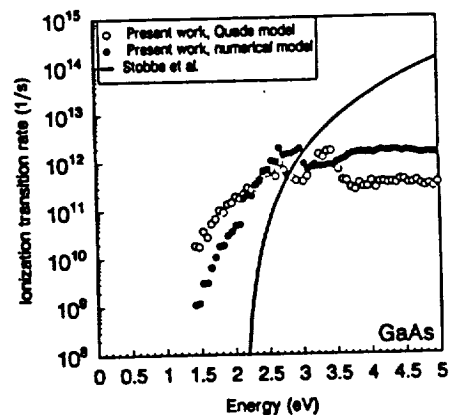


FIG. 9. Calculated impact-ionization transition rate in gallium arsenide as a function of electron energy measured from the conduction-band minimum using the k-dependent numerical model and the Quade model. The energy-dependent rate is calculated by performing a further integration of the k-dependent rate following the approach explained in Ref. 15. The energy-dependent rate obtained by Stobbe and co-workers (Ref. 21) is also included.

The results of Kane,⁴⁰ Kamakura *et al.*,¹⁹ and the recently published ionization transition rate of Cartier *et al.*⁴¹ determined from a comparison of Monte Carlo simulations and soft-x-ray photoemission spectroscopy (XPS) measurements are shown as well. The numerically calculated transition rate is almost identical with that of Cartier *et al.* up to an energy of 4 eV, and seems to be closer to the results of Kane in the high-energy region. Recently, also, Kunikiyo *et al.*¹⁸ have performed similar calculations of the silicon interband impact-ionization transition rate. Their calculations are in excellent agreement as well with the transition rate of Cartier *et al.*⁴¹ The transition rate obtained on the basis of the Quade formula agrees well with the numerical model between 2 and 4 eV, while a lower rate is predicted in the low- and high-energy regions.

The energy-dependent impact-ionization transition rate for gallium arsenide, calculated using the numerical approach of Wang and Brennan¹⁶ and the Quade formula, respectively, are shown in Fig. 9. Stobbe and co-workers² have also recently calculated the impact-ionization transition rate in bulk GaAs. Their calculations, based on the fitting formula quoted in Ref. 21, are included in Fig. 9 for comparison. Inspection of Fig. 9 shows that the maxima of the transition rate, based on both the numerical model of Wang and Brennan¹⁶ and the Quade model, occurs near 3 eV in GaAs. The maximum region is more pronounced for the Quade formulation than in the case of Wang and Brennan's numerical model. The origin of the maximum can be possibly understood within the analytical formulation, since the transition rate involving each valley can be individually evaluated. The maximum in the transition rate is due to the collision of an electron within the second conduction band with an electron in the heavy-hole band, after which the electrons occupy states either both in the Γ valley (the lower peak below 3.0 eV) or one is in the Γ valley and the other in the L valley (the higher peak close to 3.4 eV). In both cases a transition of the initiating electron from the second conduc

tion band to the first one with only a slight change of the k vector (a near-vertical transition, as explained in Ref. 42) contributes to the final values of the transition rate. Near-vertical transitions have a significantly greater rate owing to the fact that the denominator in Eq. (3) becomes very small under these conditions.⁴² This region may be somewhat unrealistically pronounced in the Quade formulation due to the fact that the value of the overlap integrals is taken as an averaged constant. The actual values of the overlap integrals in this region may be quite small thereby reducing the rate if calculated. It is much more difficult to isolate the physical origin of the maximum in the numerical model. In this case, a similar classification would be questionable, mainly due to the uncertainty in how the "valleys" are defined in the real band structure. Inspection of Fig. 9 also shows that the study of Stobbe and co-workers²¹ predicts a higher ionization transition rate from the results presented here. The cause of this discrepancy may be due to the inclusion of a q -dependent dielectric function in Stobbe and co-workers' formulation as well as differences in the energy band structure utilized in the two calculations. In Stobbe and co-workers' work the band structure is calculated using an empirical pseudopotential method in contrast to the k - p method adopted in the present work.

To further investigate the impact-ionization transition rate formulations, the energy dependence of the quantum yield was also calculated for both the numerical and Quade models. The quantum yield is defined as the average number of impact-ionization events caused by a high-energy injected electron until its kinetic energy relaxes below the ionization threshold through scattering and/or ionization events. While the impact-ionization coefficient strongly depends on the energy distribution function and therefore mainly on the inelastic electron-phonon scattering rate, the quantum yield is almost a linear function of the ratio of the ionization transition rate to the electron-phonon scattering rate.⁴¹ The quantum yield provides a useful means of assessing the relative hardness or softness of the impact-ionization threshold and enables an additional, independent evaluation of the appropriateness of the model used.

To the authors' knowledge, the only presently existing experimentally based data for the quantum yield apply to bulk silicon. Calculations of the quantum yield in silicon are presented in Fig. 10 using both the numerical and Quade models. The data of Cartier *et al.*,⁴¹ derived from a comparison of Monte Carlo simulations and XPS measurements, which were shown to be in good agreement with other measurements,⁴¹ are also plotted in Fig. 10 for comparison. While fairly good agreement over the full energy range between the model based on the numerically calculated ionization rate and the data from Ref. 41 can be seen, the Quade formula predicts a quantum yield which is substantially lower in the low-energy region. Nevertheless, the Quade formula results correspond fairly well to the other data above 2.5 eV. These results can be understood on the basis of the threshold energy for impact ionization. In the case of the Quade formula the threshold energy is well defined and is substantially higher than the value of the gap. Subsequently the ionization-transition rate in this region is lower for the

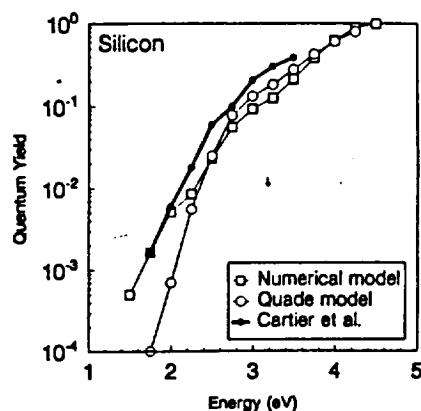


FIG. 10. Calculated quantum yield as a function of electron energy in silicon measured from the conduction-band minimum, obtained using the k -dependent numerical model and the Quade model. Results of Cartier *et al.* (Ref. 41) are also shown.

Quade model, as can be seen in Fig. 8, which results in a lower quantum yield. Conversely, in the numerical model, the effective threshold is simply the energy gap and subsequently ionization events can possibly occur at low energies in this model leading to a higher quantum yield.

Although no experimental data for bulk GaAs are presently available, it is nevertheless instructive to compare the calculated quantum yield dependencies on energy using the numerical and Quade models. The calculated quantum yield versus electron energy in GaAs using the numerical and Quade models is shown in Fig. 11. Comparison of the calculations reveals good correspondence between these two models in the energy range from 2.5 to 3.5 eV, while at higher energies the Quade model [with the particular choice of the prefactor F appearing in Eq. (7), as described above] predicts a lower quantum yield than the numerical model. At lower energies the Quade model predicts a higher quantum yield than the numerical model. This discrepancy can be understood as follows. An identical phonon scattering rate is used in both models of the ionization rate. As stated by Cartier *et al.*⁴¹ the quantum yield depends linearly upon the ratio of the impact-ionization transition rate to the phonon scattering

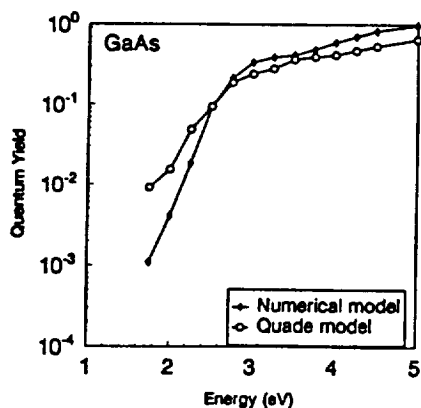


FIG. 11. Calculated quantum yield as a function of electron energy in gallium arsenide measured from the conduction-band minimum, obtained using the k -dependent numerical model and the Quade model.

rate. Subsequently, the quantum yield within each model, given that the scattering rates are identical, depends then simply upon the ratio of the transition rates and will follow the same trends as shown in Fig. 9.

The value of the overlap integrals used in the evaluation of the transition rate given by Eqs. (2) and (3) is evaluated directly in the numerical approach. As mentioned above, this requires extensive numerical computation. By approximating these overlap integrals as a constant, the factor F in Eq. (7), a new single parameter is introduced greatly reducing the numerical complexity required to evaluate the transition rate but at the expense of a new unknown parameter. In the calculations presented above, F is adjusted such that the analytical model yields satisfactory agreement with the experimental data. To check to see if the determined values of F are reasonable, it is useful to compare these values to those determined using the sum rule quoted by Ridley.¹² Using the sum rule, the squared overlap integral between the first conduction band and the heavy-hole band in gallium arsenide is determined to be 0.38. More elaborate calculations made by Burt *et al.*^{43,44} using wave functions obtained from a $k \cdot p$ calculation show that this value is $\sim 10^{-2}$, which is in fair agreement with the results presented here for GaAs, $F=0.005$ for the second conduction band. In the case of silicon, the value $F=0.24$ obtained for the impact ionization originating in both the first and second conduction bands, is higher than the value of $F \sim 0.05$,⁴⁵ or $F=0.01$, quoted in Refs. 46 and 47, obtained by theoretical evaluations. It should be noted, though, that a discrepancy between the experimentally determined recombination coefficients for Auger recombination, and the theoretically obtained ones, which also contain the same overlap integrals, was observed in bulk silicon. This discrepancy is thought to be possibly due to an underestimation of the overlap integrals by roughly an order of magnitude.⁴⁸

IV. CONCLUSIONS

In this article the electron interband impact-ionization rate calculated by an ensemble Monte Carlo simulation has been presented for both bulk silicon and gallium arsenide with the expressed purpose of comparing different formulations of the transition rate. Specifically, three different treatments of the transition rate are examined: the traditional Keldysh formula, a new k -dependent analytical formulation first derived by Quade and co-workers,²² and a more exact, numerical method of Wang and Brennan.^{15,16} The numerical method of Wang and Brennan^{15,16} contains no readily adjustable parameters and is similar to other numerical techniques that are currently attaining prominence. The numerical technique thus serves as a standard to which the Quade result has been compared. Unfortunately, the completely numerical technique, although fundamentally based, requires extensive numerical computation making it somewhat unattractive. As is well known, the Keldysh formula, although commonly used in Monte Carlo calculations of the impact-ionization rate owing to its relative simplicity and computational efficiency, suffers from several limitations as well. Chief among these is the fact that it is derived assuming parabolic energy

bands for all of the participating carriers, and includes no k dependence in the rate. Additionally, due to its reliance on two adjustable parameters, more than one set of parameters can be used within the Keldysh formula to produce satisfactory agreement with experiment. As a result, the Keldysh formula cannot by itself predict whether a material has a soft or hard ionization threshold. Therefore, there is a need to develop a more computationally efficient transition rate formulation than the direct, numerical technique, yet retain the k dependence and details of the band structure in its formulation. In this way, the physical nature of the ionization process can be preserved (soft or hard threshold, etc.) while maintaining computational efficiency. The Quade formula, although it contains one adjustable parameter corresponding physically to the squared overlap integrals of the periodic parts of the Bloch functions, includes the full details of the band structure for the initiating carrier. In this article, we have provided the first comparison of this formula to the numerical technique to determine if this approach yields qualitatively similar results while providing far better computational efficiency.

The calculations presented here show that the model based on the Quade formulation produces similar results to the completely numerical calculations for some quantities, primarily the ionization rate as a function of field. In contrast, to the Keldysh formula, where no k -vector dependence of the impact-ionization transition rate is included, the transition rate in the k -dependent numerical and analytical (Quade formula) cases is shown to be strongly dependent on the initiating electron k vector. It is also shown that both k -dependent models predict that the second conduction band dominates the ionization process in GaAs and silicon; however, the quantum yield predicted using the analytical model of Quade and co-workers²² does not agree closely with the available experimentally derived data of Cartier *et al.*⁴¹ or that predicted using the numerical model of Wang and Brennan.^{15,16} As such the Quade formula, although far more complete than the Keldysh formula, may not be a completely reliable replacement for the numerical models in many situations.

Further comparison of the Quade and numerical models reveals that they predict that the effective threshold for impact ionization is relatively soft, implying that the carriers drift to substantially high energies before suffering an ionization event. It is shown that the average energy of the electrons is high when compared to that expected for a hard threshold model. Thus, many of the important qualitative results obtained by adopting the computationally intensive numerical approach, i.e., the dominance of the second conduction band, the nature of the ionization threshold, and the importance of the k dependence, can be recovered if the analytical k -dependent formulation of Quade and co-workers²² is used. Use of the Quade formula may not always be fully justifiable, however, since it contains some *ad hoc* parametrization which cannot be ascertained from first principles and it apparently does not reproduce the quantum yield data. Additionally, it is possible that the Quade formula may overestimate the impact-ionization transition rate at those points in the k space where near-vertical transi-

tions are allowed due to the failure to include the actual values of the overlap integrals for these transitions.

ACKNOWLEDGMENTS

The authors would like to thank Dr. C. J. Summers and Dr. B. K. Wagner for many helpful discussions. This work was supported in part by the National Science Foundation through a Presidential Young Investigator Award made to K. Brennan, through Bell Northern Research Grant No. I25A3J and by ARPA through contract to NASA, NAGW-2753. Computing support was provided by the Digital Equipment Corp. through Contract No. E21-H36.

- ¹ W. Shockley, *Solid-State Electron* 2, 35 (1961).
- ² P. A. Wolff, *Phys. Rev.* 95, 1415 (1954).
- ³ G. Baraff, *Phys. Rev.* 128, 2507 (1962).
- ⁴ R. C. Curby and D. K. Ferry, *Phys. Status Solidi A* 15, 319 (1973).
- ⁵ P. A. Lebowitz and P. J. Price, *Solid State Commun.* 9, 1221 (1971).
- ⁶ H. Shichijo and K. Hess, *Phys. Rev. B* 23, 4197 (1981).
- ⁷ M. V. Fischetti and S. E. Laux, *Phys. Rev. B* 38, 9721 (1988).
- ⁸ C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer, Vienna, 1989).
- ⁹ J. Bude, K. Hess, and G. J. Iafrate, *Phys. Rev. B* 45, 10 958 (1992).
- ¹⁰ W. Quade, F. Rossi, and C. Jacoboni, *Semicond. Sci. Technol.* 7, B502 (1992).
- ¹¹ P. D. Yoder, J. M. Higman, J. Bude, and K. Hess, *Semicond. Sci. Technol.* 7, B357 (1992).
- ¹² B. K. Ridley, *Quantum Processes in Semiconductors*, 2nd. ed. (Oxford University Press, Oxford, 1988).
- ¹³ J. Y. Tang and K. Hess, *J. Appl. Phys.* 54, 5139 (1983).
- ¹⁴ L. V. Keldysh, *Zh. Eksp. Teor. Fiz.* 48, 1962 (1965) [*Sov. Phys. JETP* 21, 1135 (1965)].
- ¹⁵ Y. Wang and K. F. Brennan, *J. Appl. Phys.* 75, 313 (1994).
- ¹⁶ Y. Wang and K. F. Brennan, *J. Appl. Phys.* 76, 974 (1994).
- ¹⁷ N. Sano and A. Yoshii, *Phys. Rev. B* 45, 4171 (1992).
- ¹⁸ T. Kunikiyo, M. Takenaka, Y. Kamakura, M. Yamaji, H. Mizuno, M. Morifuji, K. Taniguchi, and C. Hamaguchi, *J. Appl. Phys.* 75, 297 (1994).
- ¹⁹ Y. Kamakura, H. Mizuno, M. Yamaji, M. Morifuji, K. Taniguchi, C. Hamaguchi, T. Kunikiyo, and M. Takenaka, *J. Appl. Phys.* 75, 3500 (1994).
- ²⁰ M. Stobbe, A. Konies, R. Redmer, J. Henk, and W. Schattke, *Phys. Rev. B* 44, 11 105 (1991).
- ²¹ M. Stobbe, R. Redmer, and W. Schattke, *Phys. Rev. B* 49, 4494 (1994).
- ²² W. Quade, E. Scholl, and M. Rudan, *Solid-State Electron.* 36, 1493 (1993).
- ²³ Y. C. Chang, D. Z.-Y. Ting, J. Y. Tang, and K. Hess, *Appl. Phys. Lett.* 42, 76 (1983).
- ²⁴ W. Fawcett, A. D. Boardman, and S. Swain, *J. Phys. Chem. Solids* 31, 1963 (1970).
- ²⁵ J. Y. Tang, Ph.D. thesis, University of Illinois, Urbana, IL, 1982.
- ²⁶ Y. Ohno, *J. Appl. Phys.* 64, 4549 (1988).
- ²⁷ C. Jacoboni, C. Canali, G. Ottaviani, and A. A. Quaranta, *Solid-State Electron.* 20, 77 (1977).
- ²⁸ K. F. Brennan, D. H. Park, K. Hess, and M. A. Littlejohn, *J. Appl. Phys.* 63, 5004 (1988).
- ²⁹ K. F. Brennan, N. Mansour, and Y. Wang, *Comput. Phys. Commun.* 67, 73 (1991).
- ³⁰ D. H. Park, Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA, 1989.
- ³¹ Landolt-Börnstein, *Numerical Data and Functional Relationships in Science and Technology*, edited by K.-H. Hellwege and O. Madelung (Springer, New York, 1982).
- ³² M. L. Cohen and T. K. Bergstresser, *Phys. Rev.* 141, 789 (1966).
- ³³ R. van Overstraeten and H. DeMan, *Solid-State Electron.* 13, 583 (1970).
- ³⁴ W. N. Grant, *Solid-State Electron.* 16, 1189 (1973).
- ³⁵ M. H. Woods, W. C. Johnson, and M. A. Lambert, *Solid-State Electron.* 16, 381 (1973).
- ³⁶ G. E. Bulman, V. M. Robbins, K. F. Brennan, K. Hess, and G. E. Stillman, *IEEE Electron Device Lett.* EDL-4, 181 (1983).
- ³⁷ N. Sano, T. Aoki, and A. Yoshii, *Appl. Phys. Lett.* 55, 1418 (1989).
- ³⁸ N. Sano, M. Tomizawa, and A. Yoshii, *Appl. Phys. Lett.* 56, 653 (1990).
- ³⁹ J. Bude and K. Hess, *J. Appl. Phys.* 72, 3554 (1992).
- ⁴⁰ E. O. Kane, *Phys. Rev.* 159, 624 (1967).
- ⁴¹ E. Cartier, M. V. Fischetti, E. A. Eklund, and F. R. McFeely, *Appl. Phys. Lett.* 62, 3339 (1993).
- ⁴² Y. Wang and K. F. Brennan, *J. Appl. Phys.* 71, 2736 (1992).
- ⁴³ M. G. Burt and C. Smith, *J. Phys. C* 17, L47 (1984).
- ⁴⁴ M. G. Burt, S. Brand, C. Smith, and R. A. Abram, *J. Phys. C* 17, 6385 (1985).
- ⁴⁵ J. Hult, *Phys. Status Solidi A* 8, 173 (1971).
- ⁴⁶ D. Hill and P. T. Landesberg, *Proc. R. Soc. London Ser. A* 347, 547 (1976).
- ⁴⁷ D. Hill, *Proc. R. Soc. London, Ser. A* 347, 565 (1976).
- ⁴⁸ J. Hult, in 1972 International Conference on Semiconductor Physics, Warsaw, 1972, p. 1097.

Design Issues of GaAs and AlGaAs Delta-Doped p-i-n Quantum-Well APD's

Yang Wang, *Member, IEEE*, and Kevin F. Brennan

Abstract—We examine the basic design issues in the optimization of GaAs delta-doped and AlGaAs delta-doped quantum-well avalanche photodiode (APD) structures using a theoretical analysis based on an ensemble Monte Carlo simulation. The devices are variations of the p-i-n doped quantum-well structure previously described in the literature. They have the same low-noise, high-gain and high-bandwidth features as the p-i-n doped quantum-well device. However, the use of delta doping provides far greater control of the doping concentrations within each stage possibly enhancing the extent to which the device can be depleted. As a result, it is expected that the proposed devices will operate at higher gain levels (at very low noise) than devices previously developed.

I. INTRODUCTION

AVALANCHE photodiodes (APD's) made from compound semiconductors (such as GaAs, InP, etc.) have attracted much attention for light detection because they are spectrally well matched to the wavelengths of interest in communications and imaging systems. For high-gain, high-speed, and low-noise operation, it is very important that avalanching devices exhibit single-carrier-initiated, single-carrier-multiplication (SCISM) [1]. However, most bulk compound semiconductors have nearly equal electron and hole ionization coefficients [2]. Hence, APD's made from these bulk semiconductor compounds exhibit poor noise performance [3].

Chin *et al.* [4] first suggested that the electron ionization rate can be selectively enhanced over the hole ionization rate by using a multiquantum-well structure. In such devices, layers of narrow bandgap material are sandwiched between layers of wide bandgap material. Due to the nonlinear dependence of the ionization rate on the electric field, the ionization rate may be enhanced over the corresponding bulk rate in such a device [5]. Capasso later recognized that this concept of "bandgap engineering" offers a new dimension to semiconductor device design [6], which has led to the suggestion of many new device concepts. Specifically, in addition to the simple multiquantum-well APD first introduced, many alternative APD device structures have been devised.

Among the various device structures introduced, doped quantum-well structures [7], [8] seem to be the most promis-

ing. The basic structure of the doped quantum-well devices consists of alternating layers of wide and narrow bandgap material, with a p-i-n region formed within the wide bandgap layer. Each unit cell contains five separate layers. These are p-i-n doped regions within the wide bandgap material, typically AlGaAs, followed by a nearly intrinsic relatively small bandgap material, typically GaAs, and a nearly intrinsic wide bandgap, AlGaAs, layer. Application of a reverse bias to the device acts to deplete the built-in junctions, leading to an asymmetric electric field profile within each unit cell. The combined action of the built-in electric field, overall bias field, and the conduction bandedge discontinuity acts to heat the electrons to high energy upon their injection into the narrow-gap layer. It is within the narrow gap layer that most ionization events occur. The holes, on the other hand, are injected from the high field, p-i-n layers, into the wide gap, nearly intrinsic AlGaAs layer, where they cool prior to entering the small bandgap GaAs region. The hole temperature then is substantially less than the electron temperature within the GaAs layer. As a result, the electron ionization rate is very much larger than the hole ionization rate. Monte Carlo calculations indicate that these devices can possibly show approximately four orders of magnitude enhancement in the ratio of electron and hole ionization rates over the corresponding ratio for bulk materials [8]. The large enhancement of the electron-to-hole ionization rate ratio apparently stems from the asymmetric heating of the electron and hole distributions prior to their injection into the narrow gap layer.

Though the original design [7], [8] offers the potential of high gain at low noise, its experimental realization is presently difficult to reliably achieve due to the high doping concentrations that are required within the wide-gap AlGaAs layer. Currently, high n-type doping of AlGaAs is difficult to attain. In order to circumvent the practical limitation placed on the doping concentrations, an alternative design was suggested by Brennan *et al.* [9]. In this design, high doping within the wide-bandgap material is avoided by forming the p-i-n layer within the narrow-gap region. With the proposed modification, the p-i-n doped quantum-well device is predicted to again show a substantial enhancement of the electron-to-hole ionization rate ratio. As such, low-noise, high-gain and large-bandwidth behavior should be attainable.

Experiments made by Aristan *et al.* [10] on GaAs based doped p-i-n devices have shown promising results, demonstrating the lowest noise performance of any compound semiconductor APD ever measured to date. However, their devices

Manuscript received January 20, 1993; revised August 3, 1993. This work was sponsored in part by DARPA and NASA through contract to NASA, NAGW-2753. Additional support was received by the National Science Foundation through a Presidential Young Investigator Award made to K. Brennan.

The authors are with the School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0250 USA.

IEEE Log Number 9216386.

failed to exhibit low-noise performance at gains greater than 5. It is speculated that the increased noise at high gain levels arises from the failure to fully deplete all of the stages of the device due to an imbalance in the n and p type doping concentrations present within each stage [11]. It is believed that only some of the stages within the device then become depleted by the reverse bias. As a result, a further increase of the reverse bias increases the overall field within the device. Subsequently, the breakdown at higher bias is due predominately to the action of the overall electric field and not from the built-in p-i-n layers. As a result, the hole ionization rate begins to become important, leading to a significant increase in the noise of the device, as has been observed. Due to the fact that the doping concentrations are extremely high within the original doped p-i-n APD designs, even small variations in the doping concentrations can lead to significant charge imbalance. It is desirable, then, to develop a design in which control of the doping concentration can be enhanced enabling the full depletion of the active region. If all of the stages of the device can be fully depleted, then electron ionization should occur within each stage, greatly increasing the gain of the overall device.

In this paper we present a new variation of the original doped p-i-n APD devices. In the new structures, the built-in p-i-n junctions are formed through the use of delta doping [12] instead of volume doping. Single sheets of dopants are selectively added to form p-i-n junctions within either the AlGaAs or GaAs layers. It is expected that the device can be more readily depleted since charge balance within each stage is more easily achieved. Through a judicious choice of the doping concentration within each sheet, and the other device parameters, low-noise operation at higher gain than measured before should be possible. Calculations based on a many-particle ensemble Monte Carlo simulation are presented here to illustrate the basic design criteria and to evaluate the potential of the delta-doped APD device. In Section II, the device structure and a brief review of the modeling procedure are discussed. The calculated results are discussed in Section III. Finally, conclusions are drawn in Section IV.

II. DEVICE STRUCTURE AND MODEL DESCRIPTION

The first device structure proposed, which is a modification of the narrow-bandgap-material-doped (in our case, GaAs-doped) quantum-well structure consists of repeated unit cells of AlGaAs and delta-doped p-i-n GaAs layers as shown in Fig. 1. The second device structure is similar to the first one, but the doping layers are placed within a wide-bandgap material (AlGaAs-doped). Therefore, the two designs are similar except for the fact that in the second structure, the doping layers and intrinsic layer sandwiched between them are formed in AlGaAs rather than GaAs, as shown in Fig. 1. The top layer of the structures can be formed of either p^+ GaAs or p^+ $\text{Al}_x\text{Ga}_{1-x}\text{As}$, depending upon the spectral requirements. If the top layer is chosen to be GaAs, an intrinsic layer between the absorption layer, and the intrinsic AlGaAs layer is needed to avoid electron trapping at the first heterostructure interface.

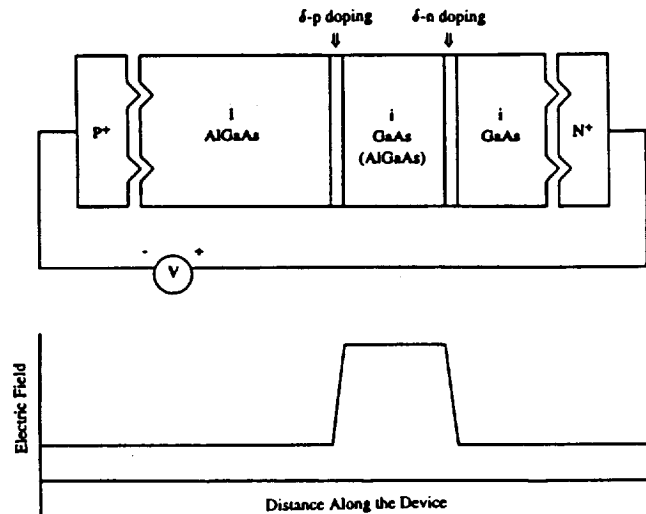


Fig. 1. Sketch of the basic unit cell of the device, along with the electric field profile.

For the GaAs-doped structure, the basic unit cell has five layers made of two materials with different bandgaps. The p- and n-doped layers are formed in the narrow-bandgap material, GaAs, and are composed of about only one monolayer. As in the original design [9], the electrons move from left to right in the diagram, while the holes move in the opposite direction. The electrons traverse into the high-field region from the widegap material, and are heated by the combined action of the built-in field of the p-i-n layers and the overall applied bias field prior to the injection into the low-field, narrow-bandgap layer. Conversely, after being heated by the built-in electric field, the holes are injected into the wide bandgap layer, where the ionization threshold is much higher. Hence, few if any hole ionization events will occur. If the high-field layer width is chosen in such a way that the holes are not heated sufficiently to reach ionization threshold, then the hole ionization can be fully suppressed. At the same time, a substantial electron ionization rate may be retained, because the electrons reach their peak energy within the narrow-gap, GaAs layer. The key issues in design of this device are that the high-field region must be sufficiently narrow, and the built-in field be sufficiently low so that the holes do not impact ionize within the GaAs, yet significant electron ionization is still retained.

In the case of the AlGaAs doped structure, the doped layers are formed in the wide-bandgap material, AlGaAs. This variation of the device is often called the doped barrier device. The electrons are heated in the high-field region in the AlGaAs layer prior to injection into the GaAs layer. Though the key issues in design of this device are still that the high-field region must be narrow enough and the built-in field be sufficiently low to suppress hole ionization, the conditions are less restrictive in this structure than that of the GaAs-doped structure, due to the fact the ionization threshold is higher in the AlGaAs layer.

The analysis of the device presented here is based on an ensemble Monte Carlo simulation incorporating the full details of the band structures, all the relevant phonon scattering

TABLE I
EFFECT OF VARIATIONS OF LAYER WIDTHS, BUILT-IN FIELD STRENGTH, AND APPLIED ELECTRIC FIELD STRENGTH
ON THE CALCULATED ELECTRON AND HOLE IONIZATION COEFFICIENTS IN A GaAs-DELTA-DOPED MQW STRUCTURE

Device Structure						
Low-Field GaAs Layer Width = 200 Å Applied Field = 75 kV/cm						
Built-in Field = 500 kV/cm AlGaAs Layer Width = 500 Å						
High-Field GaAs Layer Width = 200 Å						
Case 1: Varying the High-Field Layer Width						
High Field GaAs Layer Width (Å)	Built-in Field (kV/cm)	α (cm ⁻¹)	P	β (cm ⁻¹)	Q	% of Holes Ionizing
200.0	500.0	7.45×10^3	0.095	0.0	0.0	0.0
225.0	500.0	1.06×10^4	0.125	3.23×10^1	2.97×10^{-4}	0.22
250.0	500.00	1.45×10^4	0.173	9.7×10^1	9.2×10^{-4}	0.67
Case 2: Varying the Built-in Field						
High Field GaAs Layer Width (Å)	Built-in Field (kV/cm)	α (cm ⁻¹)	P	β (cm ⁻¹)	Q	% of Holes Ionizing
200.0	500.0	7.45×10^3	0.095	0.0	0.0	0.0
200.0	525.0	1.03×10^4	0.123	1.6×10^1	1.45×10^{-4}	0.11
200.0	550.00	1.29×10^4	0.155	3.4×10^1	3.0×10^{-4}	0.22
Case 3: Varying the Applied Electric Field, (High-Field Layer Width = 200.0 Å)						
Applied Field (kV/cm)	Built-in Field (kV/cm)	α (cm ⁻¹)	P	β (cm ⁻¹)	Q	% of Holes Ionizing
75.0	500.0	7.45×10^3	0.095	0.0	0.0	0.0
85.0	500.0	9.65×10^3	0.113	3.1×10^1	2.94×10^{-4}	0.22
100.0	500.00	1.30×10^4	0.134	5.73×10^1	5.2×10^{-4}	0.44
Case 4: Varying the Low-Field AlGaAs Layer Width (Low-Field GaAs Width = 150 Å)						
Low-Field AlGaAs Layer Width (Å)	Built-in Field (kV/cm)	α (cm ⁻¹)	P	β (cm ⁻¹)	Q	% of Holes Ionizing
500.0	500.0	8.31×10^3	0.078	0.0	0.0	0.0
450.0	500.0	8.0×10^3	0.076	3.1×10^1	2.5×10^{-4}	0.22
400.0	500.00	8.7×10^3	0.075	7.5×10^1	5.6×10^{-4}	0.56

mechanisms, and the Keldysh formula [8] for the impact ionization transition rate using a moderately soft threshold energy. The full details of our approach have been reported exhaustively in the literature [8], [9], [13] and will not be repeated here. Through computer experiments, the effect of the device geometry and electric fields on the device performance can be isolated and independently assessed.

III. CALCULATED RESULTS

In general, the ionization rates depend upon several independent device parameters, such as the layer widths, doping concentrations (i.e., the built-in field strength), and the applied electric field strength. This dependence can be most easily studied through computer experiments. One parameter at a time is varied, and its effect on the ionization rates and other quantities of interest, such as trapping, is determined. In this way, the optimal design of the device can be reached.

There are five independent parameters that can be adjusted in the computer experiments: the applied electric field, the built-in electric field, which is related to the charge density by the equation $E_b = qd/\epsilon$ [12], the high-field layer width, the low-field GaAs layer width, and the low-field AlGaAs layer width. The high-field layer width is defined as the

layer sandwiched between the two delta-doped regions, as shown in Fig. 1. The low-field AlGaAs layer is the first layer immediately to the right of the p+ contact, as shown in Fig. 1. The low-field GaAs layer is the layer placed immediately to the right of the n-type delta-doped region. There are two major issues that govern the device performance. First, the electron temperature must be high within the GaAs layer to obtain a high impact ionization rate. Second, the hole temperature within the GaAs layer must be low enough to inhibit hole ionization. In the following we will examine how each individual device parameter influences the two design issues.

Table I shows the effects of the variation of different device parameters on the electron and hole impact ionization probabilities per stage, P and Q respectively, for the GaAs doped device. The electron and hole ionization rates as a function of inverse distance, α and β , are also reported. Because the hole ionization rate is designed to be and is calculated to be small for the designs considered, we report the percentage of holes that ionize in addition to Q and β . The device parameters and electric field strengths are listed at the top of the table. For each case listed, one parameter is varied at a time keeping the other parameters fixed at their original listed

TABLE II
EFFECT OF VARIATIONS OF LAYER WIDTHS, BUILT-IN FIELD STRENGTH, AND APPLIED ELECTRIC FIELD STRENGTH ON
THE CALCULATED ELECTRON AND HOLE IONIZATION COEFFICIENTS IN AN AlGaAs-DELTA-DOPED MQW STRUCTURE

Device Structure						
Low-Field GaAs Layer Width = 100 Å Applied Field = 175 kV/cm						
Built-in Field = 400 kV/cm AlGaAs Layer Width = 500 Å						
High-Field GaAs Layer Width = 150 Å						
Case 1: Varying the High-Field Layer Width						
High Field AlGaAs Layer Width (Å)	Built-in Field (kV/cm)	α (cm ⁻¹)	P	β (cm ⁻¹)	Q	% of Holes Ionizing
150.0	400.0	1.47×10^4	0.110	0.0	0.0	0.0
175.0	400.0	2.04×10^4	0.159	1.39×10^1	1.08×10^{-4}	0.11
200.0	400.00	2.42×10^4	0.194	5.64×10^1	4.51×10^{-4}	0.44
Case 2: Varying the Built-in Field						
High Field AlGaAs Layer Width (Å)	Built-in Field (kV/cm)	α (cm ⁻¹)	P	β (cm ⁻¹)	Q	% of Holes Ionizing
150.0	400.0	1.47×10^4	0.110	0.0	0.0	0.0
150.0	425.0	1.75×10^4	0.132	2.83×10^1	2.12×10^{-4}	0.22
150.0	450.00	2.14×10^4	0.160	2.81×10^1	2.11×10^{-4}	0.22
Case 3: Varying the Applied Electric Field (High-Field Layer Width = 200.0 Å)						
Applied Field (kV/cm)	Built-in Field (kV/cm)	α (cm ⁻¹)	P	β (cm ⁻¹)	Q	% of Holes Ionizing
175.0	400.0	1.47×10^4	0.110	0.0	0.0	0.0
185.0	400.0	1.69×10^4	0.126	4.0×10^1	3.0×10^{-4}	0.33
200.0	400.00	2.08×10^4	0.154	7.73×10^1	5.8×10^{-4}	0.67
Case 4: Varying the Low-Field AlGaAs Layer Width						
Low-Field AlGaAs Layer Width (Å)	Built-in Field (kV/cm)	α (cm ⁻¹)	P	β (cm ⁻¹)	Q	% of Holes Ionizing
500.0	400.0	1.47×10^4	0.110	0.0	0.0	0.0
450.0	400.0	1.55×10^4	0.108	1.42×10^1	1.00×10^{-4}	0.11
400.0	400.0	1.66×10^4	0.109	2.96×10^1	1.92×10^{-4}	0.22

values. It is obvious from the table that the electron ionization rate is most sensitive to the high-field GaAs layer width, the applied electric field, and the built-in electric field strength. The electron ionization rate is essentially invariant with respect to fluctuations in the low-field AlGaAs layer width, at least for the values considered here. The percentage change in the electron ionization rate as a function of percentage change in each parameter is greatest for changes in the built-in field. This is also true for the holes, though the hole ionization rate depends nearly as much on the high-field layer width and the low-field AlGaAs layer width. Subsequently, the device noise and gain performance are most sensitive to fluctuations in the doping concentration. Nevertheless, fluctuations in the overall bias field and the high-field GaAs layer width can also be important.

Table II shows the effects of the variation of different device parameters on the electron and hole impact ionization probabilities per stage, P and Q respectively, for the AlGaAs-doped device. As in the case of the GaAs-doped structure, both the electron and hole ionization rate are very sensitive to the high-field layer width, the applied electric field, and the built-in electric field strength. However, the difference here

is that the percentage change in the hole ionization rate as a function of percentage change in each parameter is greatest for changes in the applied electric field. This is because, first that the high-field region lies within the wide-bandgap material, where the ionization threshold is higher. Second, the applied field used in the simulation is relatively high, chosen so as to avoid hole trapping at the heterointerface. The applied field alone is close to that needed to cause hole ionization within GaAs. Therefore, a small increase in the applied field will lead to a relatively large increase in the hole ionization rate in the device.

The hole temperature in the GaAs low-field layers not only depends on the high-field layer width and built-in field, but also depends on the AlGaAs layer width, as shown by Case 4 in both Tables I and II. The holes must cool within the AlGaAs to a temperature low enough that upon reinjection into the GaAs layers they do not subsequently obtain the energy necessary to impact ionize. If the AlGaAs layer is made too narrow, the holes reenter the GaAs with energy sufficiently high to reach the ionization threshold in GaAs. On the other hand, if the AlGaAs layer is too wide, then the bandwidth of the device will suffer.

Another important issue that needs to be taken into account is carrier trapping. If some of the carriers are trapped within the narrow-bandgap layer and are subsequently unable to be collected, the resulting device performance is poor. In the GaAs-doped structure, since holes enter the wide-bandgap material layer after being accelerated in the high-field GaAs layer, hole trapping at the heterointerface is negligible. This is of course due to the fact that the holes are heated to high energy from the action of the built-in field just prior to their encounter with the heterointerface. However, the electrons enter the wide-bandgap layer after being cooled within the low-field GaAs layer. Thus, some electrons may become trapped at the heterointerface. In order to explore the effect of carrier trapping in the device, we track the electron locations within the device and report those locations after 14 ps of simulation time. We choose 14 ps of simulation time since the system has reached steady-state conditions at this point. Fig. 2 shows electron position profiles after 14 ps of simulation time for three device structures, varying only in the width of the low-field GaAs region. In Fig. 2(a), there is no significant electron trapping, as evidenced by the lack of counts in the left tail of the distribution. In contrast, as the low-field GaAs layer width increases in size, some of the electrons become trapped in the initial stages of the device, as can be seen from the appearance of counts in the left tail of the distribution in Fig. 2(b). Finally, at a low-field GaAs layer width of 200 Å, as shown in Fig. 2(c), electrons are trapped throughout the device, as evidenced by the counts everywhere within the structure. From these calculations, it is obvious that the low-field GaAs layer width needs to be narrow enough such that the electrons still retain sufficient energy to overcome the barrier formed by the conduction band-edge discontinuity and hence avoid trapping. On the other hand, the low-field GaAs width should be longer than the mean free distance for electron impact ionization. In addition, the GaAs layer width should be long enough such that the injected holes can relax the energy gained from the valence band-edge discontinuity in order to avoid hole multiplication.

In the AlGaAs-doped design, however, electron trapping is less significant than hole trapping. In this case, the heterojunction is formed at the interface between the low-field GaAs layer and the delta-doped AlGaAs layer. Since the holes enter the GaAs layer already cooled from their transit across the low-field AlGaAs layer, when they reach the heterointerface, if the applied field is too low, hole trapping can occur. Fig. 3 shows hole position profiles after 14 ps of simulation time for three AlGaAs-doped device structures, varying only in the width of the low-field GaAs region. As can be seen from Fig. 3(a), there is no significant hole trapping in a design with only a 50-Å-wide GaAs layer. When the GaAs layer is widened, as shown in Fig. 3(b), some of the holes become trapped in the initial stages of the device. Finally, there is serious hole trapping in Fig. 3(c) when the GaAs layer is increased to 200 Å in width. As in the case of the GaAs-doped structure, the low-field GaAs layer width needs to be narrow enough so the holes don't lose too much energy gained from the valence band edge discontinuity. Consequently, they can then overcome the barrier and avoid trapping. On the other

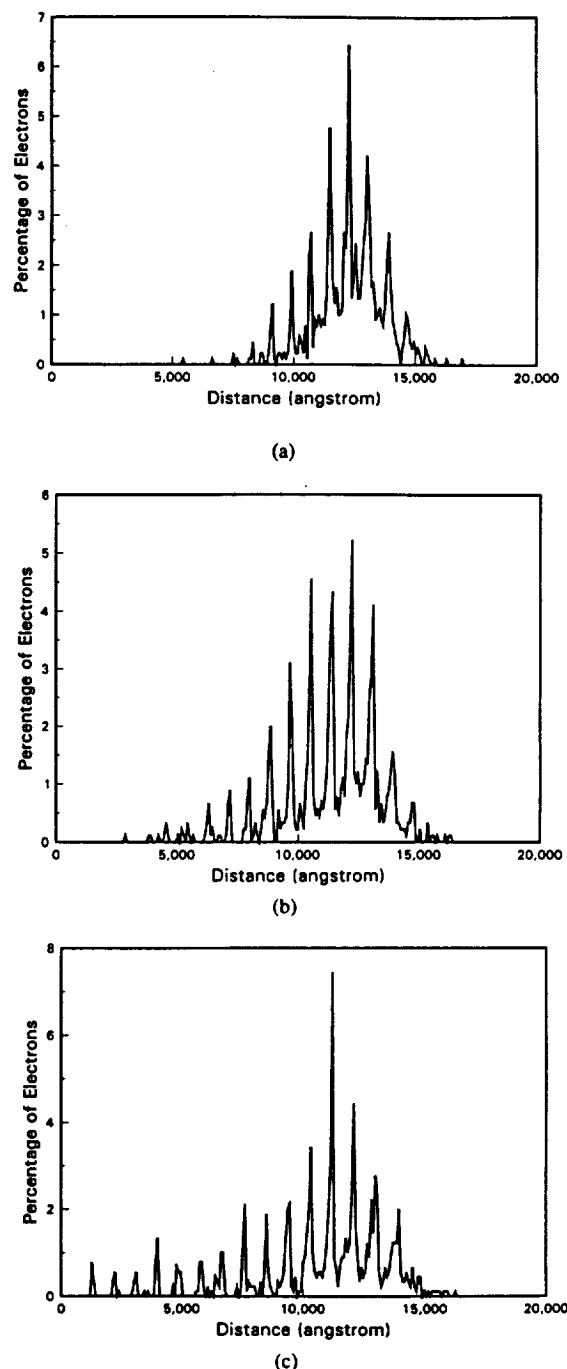


Fig. 2. Electron position profiles after 14 ps simulation time. (a) Low-field GaAs layer width = 100 Å. (b) Low-field GaAs layer width = 150 Å. (c) Low-field GaAs layer width = 200 Å. Other device parameters are: applied field = 75 kV/cm, built-in field = 500 kV/cm, i-GaAs high-field width = 200 Å, i-AlGaAs width = 500 Å.

hand, the gain per stage suffers substantially as the GaAs layer width is made smaller. This is due to the emergence of spatial quantization effects, which act to increase the effective impact ionization threshold energy as well as reduce the effective kinetic energy boost from the conduction and valence band edge discontinuities [14]. When spatial quantization within the GaAs well is appreciable, the electrons and holes can no longer thermalize to the conduction and valence band edges. Instead they can thermalize at most to the $n = 1$ states within the

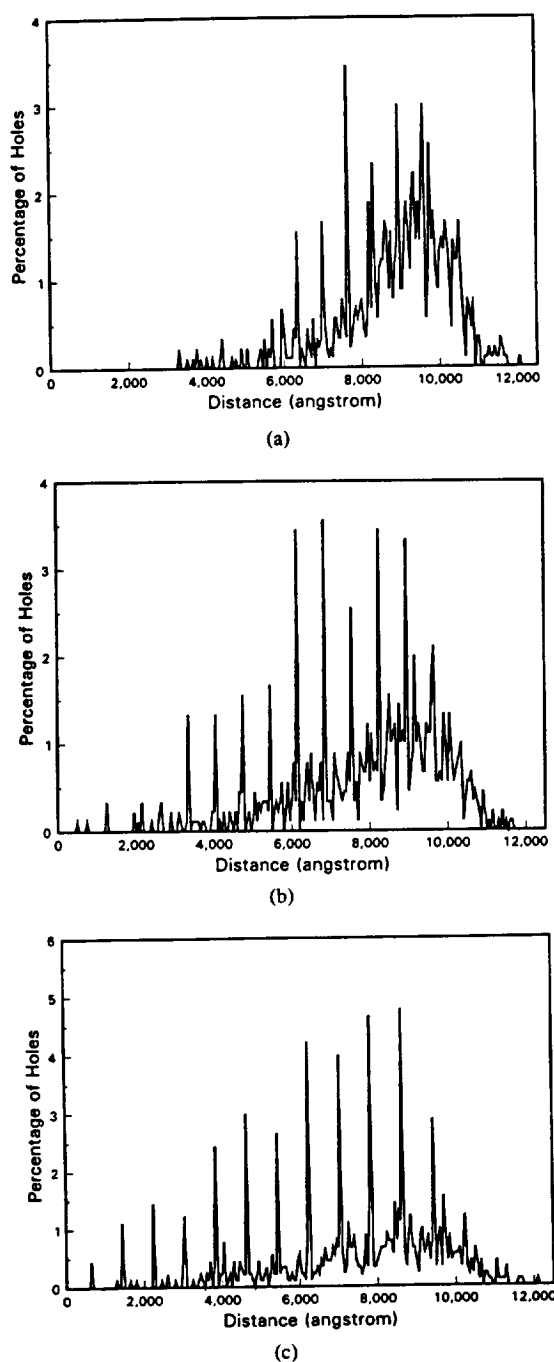


Fig. 3. Hole position profiles after 14 ps simulation time. (a) Low-field GaAs layer width = 50 Å. (b) Low-field GaAs layer width = 100 Å. (c) Low-field GaAs layer width = 200 Å. Other device parameters are: applied field = 175 kV/cm, built-in field = 400 kV/cm, i-AlGaAs high-field width = 150 Å, i-AlGaAs width = 450 Å.

well. As a result, the effective bandgap increases leading to a decrease in the impact ionization rate.

It should be noted that the trapping calculation made using the Monte Carlo simulation serves only as an estimate of the degree of trapping present in the device. The inclusion of quantum mechanical effects at the interface may reduce the carrier trapping problem, since spatial quantization levels form in which the carriers accumulate. As a result, the carriers do not thermalize to the band edge, resulting in a reduced effective barrier height. In addition, other quantum mechanical effects

such as tunneling or tunnel-assisted thermionic emission can reduce carrier trapping at the heterobarriers. The Monte Carlo calculation then serves as a "worst case" estimate of the carrier trapping. A more accurate model that includes quantum mechanical effects, such as that described by [15], may be more useful in determining the extent of trapping in these structures.

Charge trapping at heterostructure interfaces, particularly for InGaAs-InP separate absorption-multiplication APD's, has been shown to be reduced by three different schemes. These are insertion of a quaternary layer [16], insertion of a graded band-gap region [17] and the use of a doped interface dipole [18]. Miyoshi *et al.* [15] have shown that hole pile-up at the heterostructure interface is most effectively reduced using a graded barrier scheme. A graded AlGaAs barrier region could potentially be incorporated into the delta-doped APD device structure, though at the expense of added complexity.

IV. CONCLUSIONS

We have presented two new delta-doped APD devices that retain the potential performance of the original doped p-i-n quantum-well structures, yet offer a greater probability of being experimentally optimized. The primary advantage of the delta-doped designs is the greater level of control of the doping concentrations within each stage of the device afforded by the delta-doping technique. By carefully controlling the doping concentrations, full depletion of the device should be possible, thereby sustaining low-noise operation at high gain. The device stages consist of a widegap $\text{Al}_{0.35}\text{Ga}_{0.65}\text{As}$ layer followed by delta-doped p and n layers. In the first design, the doping layers are made within GaAs, while they lie within $\text{Al}_{0.35}\text{Ga}_{0.65}\text{As}$ in the second design. Using an ensemble Monte Carlo simulation, the effects of each individual device parameter, such as the layer widths, and the built-in and applied electric field strengths on the device performance are assessed.

The fundamental tradeoff present in the design of the device is that the doping concentration and high-field layer widths must be sufficiently low such that the hole ionization is completely suppressed, yet large enough to cause sizable electron ionization. Due to the different electron and hole ionization mean free paths and the asymmetry of the basic unit cell of the device, both constraints may be satisfied. The width of the low-field AlGaAs layer can also be adjusted to lower the hole ionization rate, and thus serve as an additional means of controlling the device performance. For the GaAs-doped device, the gain per stage is typically smaller under conditions of zero hole ionization. Nevertheless, this design is probably more practical since high doping is easier to achieve presently in GaAs than in AlGaAs. The AlGaAs-doped structure can be operated at higher fields yielding higher gain per stage, due to the fact that the high-field region is within the wide-bandgap material layer where the ionization threshold is higher.

Because of the asymmetry of the device structures, hole trapping is negligible in the GaAs-delta-doped structure, while electron trapping is negligible in the AlGaAs delta-doped device. However, carrier trapping can be severe if the applied

field is too low and the GaAs low field layer width is too wide. As such it is necessary to reduce the GaAs low-field layer width to minimize carrier trapping. Therefore, several device tradeoffs must be considered when designing a delta-doped APD structure in order to achieve high gain at low noise and at low carrier trapping.

ACKNOWLEDGMENT

The authors would like to thank Dr. C. J. Summers of the Georgia Tech Research Institute for many valuable discussions.

REFERENCES

- [1] K. Matsuo, M. C. Teich, and B. E. A. Saleh, "Noise properties and time response of the staircase avalanche photodiode," *IEEE Trans. Electron Devices*, vol. ED-32, pp. 2615-2623, 1985.
- [2] R. J. McIntyre, "Multiplication noise in uniform avalanche devices," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 164-168, 1966.
- [3] F. Capasso, "Physics of avalanche photodiodes," in *Semiconductors and Semimetals*, vol. 22, part D. New York: Academic, 1985.
- [4] R. Chin, N. Holonyak, Jr., G. E. Stillman, J. Y. Tang, and K. Hess, "Impact ionization in multilayered heterojunction structures," *Electron. Lett.*, vol. 16, pp. 467-489, 1980.
- [5] K. F. Brennan, K. Hess, and F. Capasso, "Physics of the enhancement of impact ionization in multi-quantum well structures," *Appl. Phys. Lett.*, vol. 50, pp. 1897-1900, 1987.
- [6] F. Capasso, "The graded bandgap and superlattice devices by bandgap engineering," in *Semiconductors and Semimetals*, vol. 24. New York: Academic, 1987.
- [7] H. Blauvelt, S. Margalit, and A. Yariv, "Single-carrier-type dominated impact ionization in multilayer structures," *Electron. Lett.*, vol. 18, pp. 375-376, 1982.
- [8] K. Brennan, "Theory of the doped quantum well superlattice APD: A new solid state photomultiplier," *IEEE J. Quantum Electron.*, vol. QE-22, pp. 1999-2016, 1986.
- [9] K. Brennan and W. T. Vetterling, "Theory of the GaAs-Doped p-i-n quantum well APD," *IEEE Trans. Electron Devices*, vol. 36, pp. 1597-1601, 1989.
- [10] P. Aristin, A. Torabi, A. K. Garrison, H. M. Harris, and C. J. Summers, "New doped multiple-quantum-well avalanche photodiode: The doped barrier $\text{Al}_{0.35}\text{Ga}_{0.65}\text{As}/\text{GaAs}$ multiple-quantum-well avalanche photodiode," *Appl. Phys. Lett.*, vol. 60, pp. 85-87, 1992.
- [11] C. J. Summers, private conversation.
- [12] E. F. Shubert, A. Fischer, and K. Ploog, "The delta-doped field-effect transistor (δFET)," *IEEE Trans. Electron Dev.*, vol. ED-33, pp. 625-632, 1986.
- [13] K. F. Brennan, D. H. Park, and Y. Wang, "Design and comparison of advanced semiconductor devices using computer experiments: Application to APD's and HEMT's," *IEEE Trans. Electron Devices*, vol. 37, pp. 536-546, 1990.
- [14] K. Brennan, "Theoretical study of multiquantum well avalanche photodiodes made from the GaInAs/AlInAs material system," *IEEE Trans. Electron Dev.*, vol. ED-33, pp. 1502-1510, 1986.
- [15] T. Miyoshi, H. Tsuchiya, and M. Ogawa, "Quantum hole transport at the heterointerface of long wavelength avalanche photodiodes," *IEEE J. Quantum Electron.*, vol. 28, pp. 25-30, 1992.
- [16] J. C. Cambell, W. T. Tsang, G. J. Qua, and B. C. Johnson, "High-speed InP/InGaAsP/InGaAs avalanche photodiodes grown by chemical beam epitaxy," *IEEE J. Quantum Electron.*, vol. 24, pp. 496-500, 1988.
- [17] F. Capasso, H. M. Cox, A. L. Hutchinson, N. A. Olsson, and S. G. Hummel, "Pseudo-quaternary GaInAsP semiconductors: A new $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}/\text{InP}$ graded gap superlattice and its applications to avalanche photodiodes," *Appl. Phys. Lett.*, vol. 45, pp. 1193-1195, 1984.
- [18] F. Capasso, A. Y. Cho, K. Mohammed, and P. W. Foy, "Doping interface dipoles: Tunable heterojunction barrier heights and band-edge discontinuities by molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 46, pp. 664-666, 1985.



Yang Wang (S'88-M'90) was born in Shanghai, China. He received the B.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984 and the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1986 and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1990.

He is currently a research engineer at the Georgia Institute of Technology. His current research interests concern high-field high-energy transport physics in compound semiconductors and semiconductor devices, and modeling and simulation of novel semiconductor devices, such as multiquantum well avalanche photodiodes and heterojunction bipolar transistors. He has more than twenty publications in some leading scientific journals.

Dr. Wang is a member of Sigma Xi, the Scientific Research Society.



Kevin F. Brennan received the B.S. degree in physics from the Massachusetts Institute of Technology, Cambridge, in 1978 and the M.S. and Ph.D. degrees in physics and electrical engineering, respectively, from the University of Illinois, Urbana-Champaign, in 1984.

He is currently Professor of Electrical and Computer Engineering at the Georgia Institute of Technology, Atlanta. His current research interests include modeling, theory and optimization of semiconductor devices and structures. Specifically, most of his work involves the study of photodetectors, avalanche photodiodes, high-field effects in semiconductors and transport in submicrometer compound semiconductor devices.

Dr. Brennan was a recipient of a National Science Foundation Presidential Young Investigator Award.

Theoretical Study of the Effect of an AlGaAs Double Heterostructure on Metal-Semiconductor-Metal Photodetector Performance

Ali F. Salem, *Student Member, IEEE*, Arlynn W. Smith, and Kevin F. Brennan, *Senior Member, IEEE*

Abstract—The impulse and square-wave input response of different GaAs metal-semiconductor-metal photodetector (MSM) designs are theoretically examined using a two dimensional drift-diffusion numerical calculation with a thermionic-field emission boundary condition model for the heterojunctions. The rise time and the fall time of the output signal current are calculated for a simple GaAs, epitaxially grown, MSM device as well as for various double-heterostructure barrier devices. The double heterostructure devices consist of an AlGaAs layer sandwiched between the top GaAs active, absorption layer and the bottom GaAs substrate. The effect of the depth of the AlGaAs layer on the speed and responsivity of the MSM devices is examined. It is found that there is an optimal depth, at fixed applied bias, of the AlGaAs layer within the structure that provides maximum responsivity at minimal compromise in speed.

I. INTRODUCTION

METAL-semiconductor-metal (MSM) photodetectors are becoming increasingly attractive in optoelectronic communication systems, high-speed chip-to-chip connections, and high-speed sampling applications [1]. The principal advantages of MSM photodetectors that make them an excellent choice for on-chip detectors are their responsivity-bandwidth performance, compatibility with existing planar integrated circuit technologies, and relatively low voltage operation [1]. Improvement in MSM detector design structures, specifically the introduction of a GaAs/AlGaAs heterostructure, has advanced the realization of these applications [2].

The basic MSM detector structure is shown in Fig. 1. This structure consists of interdigitated metal fingers formed on the top surface of a semiconductor layer. Light incident on the top surface of the MSM structure is absorbed within the underlying semiconductor resulting in the creation of electron-hole-pairs (EHP's). The application of a bias to the metallic fingers creates an electric field within the underlying semiconductor which acts to sweep the photogenerated carriers out of the device. How fast these carriers are collected and how many of them actually survive to the contacts within a particular collection time determine the speed and the responsivity,

Manuscript received October 6, 1993; revised February 22, 1994. The review of this paper was arranged by Associate Editor P. K. Bhattacharya. This work was supported in part by Digital Equipment Corp. through Contract E21-H36, in part by the National Science Foundation through a Presidential Young Investigator Award made to K. Brennan, and in part by ARPA through NASA Contract NAGW-2753.

The authors are with the School of Electrical and Computer Engineering, and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

IEEE Log Number 9401527.

GENERAL METAL-SEMICONDUCTOR-METAL STRUCTURE

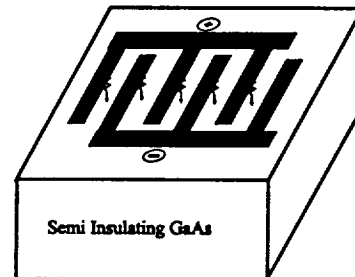


Fig. 1. Schematic diagram of an interdigitated metal-semiconductor-metal (MSM) structure.

respectively, of the detector. Carriers generated deep within the semiconductor must traverse a greater distance before they are collected at the contacts compared to those generated near the surface. Depending upon the magnitude of the electric field within the semiconductor, the time needed to collect those carriers generated deep in the device can vary drastically. Under low bias conditions, ~magnitude of 5–10 V, which is typical for most integrated circuit applications, this collection time can be prohibitively long in high speed applications.

It has been suggested that the insertion of a double heterostructure layer can improve the time response of MSM detectors [2]. The double heterostructure layer acts to block those carriers generated deep within the device structure. As a result, only those carriers photogenerated within the top absorption layer are collected leading to a fast overall response. However, as described above, the responsivity principally depends upon the number of photogenerated carriers collected at the contacts. A high responsivity, especially at low input power levels, dictates that most of the photogenerated carriers be collected. Since many of the photogenerated carriers are produced deep within the semiconductor layer, the insertion of a double heterostructure layer in order to improve the speed of the device, necessarily reduces its responsivity as well. Therefore, there exists a fundamental tradeoff between the speed of response and responsivity of a heterostructure MSM detector.

It should be noted that the absorption coefficient varies strongly with the optical wavelength. In this paper, we examine the response of the photodetector to only one incident wavelength, 840 nm. The speed of response of the detector can

be quite different at other wavelengths since the absorption may occur closer to the surface or deeper into the device. If for example, the absorption depth is much smaller than the interdigitated finger spacing then the detector's speed of response is limited by the finger separation and the placement of the AlGaAs layer has minimal effect. However, at an incident wavelength of 840 nm the absorption depth and finger separation are comparable and the placement of the AlGaAs layer becomes important. Therefore, the tradeoff between the speed of response and the responsivity discussed herein applies principally to incident light at 840 nm.

In this paper, we present results of a theoretical study of the performance of a GaAs based MSM detector examining the effect of including a sandwiched AlGaAs layer between the active GaAs layer and the substrate. A two-dimensional drift-diffusion numerical calculation in combination with a thermionic-field-emission model is employed in this study. The drift-diffusion/thermionic emission model along with the details of the numerical solution method used are described in Section II. Calibration of the model and comparison to existing models is discussed in Section III. In Section IV different GaAs MSM double heterostructure devices are examined using the model. In Section IV we examine the effect of the AlGaAs layer on the rise time, fall time, and the magnitude of the output signal current and optimize the placement of such a layer as it relates to these parameters.

II. MODEL DESCRIPTION

To understand the behavior of photo-generated carriers under the influence of drift and diffusion forces, Poisson's equation, the current continuity equations for electron and holes, and a rate equation for charged traps are solved on a two dimensional simulation domain for the electrostatic potential, and the carrier concentrations. The drift-diffusion approach is sufficient to characterize the operation of MSM detectors since these devices are typically of the order of microns in length and width. In this study, the detectors examined are restricted in dimensions to several microns where the drift-diffusion approximation is valid. The basic equations used in the simulator are [3], [4]:

$$\epsilon_0 \epsilon_s \nabla^2 \psi = -q(p - n + N_D - N_A + n_t) \quad (1)$$

$$-\frac{1}{q} \nabla \cdot J_n - G_n + R_n + \frac{\partial n}{\partial t} = 0 \quad (2)$$

$$\frac{1}{q} \nabla \cdot J_p - G_p + R_p + \frac{\partial p}{\partial t} = 0 \quad (3)$$

$$J_n = q\mu_n n E_n + qD_n \nabla n \quad (4)$$

$$J_p = q\mu_p p E_p - qD_p \nabla p \quad (5)$$

$$\frac{dn_t}{dt} = (R_n - G_n)_{SRH} - (R_p - G_p)_{SRH} \quad (6)$$

where Ψ is the electrostatic potential, n and p are the electron and hole carrier concentrations, $N_{D,A}$ are the donor/acceptor concentrations, n_t is the electron-filled trap concentration, $J_{n,p}$ are the electron/hole currents, $R_{n,p}$ and $G_{n,p}$ are the electron/hole recombination and generation rates, $\mu_{n,p}$ and $D_{n,p}$ are electron/hole mobilities and diffusivities, and the subscript SRH in (6) denotes the Shockley-Read-Hall events.

In most situations, the trap density is such that the effects of (6) are negligible [3]. The terms $R_{n,p}$ used in (2) and (3) represent the sum of Shockley-Read-Hall, radiative, and Auger recombinations for electrons and holes [4]. The optical generation rate, G , is generally assumed to be gaussian for the impulse response and is typically given as [5]

$$G_n = G_p = \frac{W_p}{h\nu} \alpha_{abs} \cdot \exp\left(-\frac{(t - t_o)^2}{\sigma_t^2}\right) \exp\left(-\frac{(x - x_o)^2}{\sigma_x^2}\right) \cdot \exp(-\alpha_{abs}y) \quad (7)$$

where W_p is the peak optical power density incident on the surface, $h\nu$ is the photon energy, α_{abs} is the absorption coefficient, x_o is the location of the center of the beam, t_o is the location of the optical input peak, σ_t and σ_x are related to the full width, half maximum (FWHM) of the pulse. In this paper, we examine the response of only a representative unit cell of the device in which the illumination is practically uniform. The beam is assumed to be centered within the unit cell. Given the dimensions of the unit cell, the lateral decay of the beam is negligible within the calculations presented here. Therefore, for simplicity, uniform illumination is assumed laterally within the unit cell and the spatial exponential term in (7) is set to unity. The metallic fingers are assumed to be completely transparent. In practice, there are of course shadowing and reflective losses at the metallic surface which would lead to a reduction in the amount of photogenerated carriers within the underlying semiconductor material. For simplicity and due to a lack of detailed information about the extent of these losses, we assume here that no losses occur.

A standard field-dependent mobility for the electron mobility, μ_n [4] is used,

$$\mu_n = \frac{\mu_{no} + v_s \frac{F^3}{F_o^4}}{1 + \left(\frac{F}{F_o}\right)^4} \quad (8)$$

where μ_{no} is the zero field mobility, F_o is the critical electric field, F is the local electric field, and v_s is the saturation velocity. The hole mobility is assumed to be constant, independent of the electric field. It is further assumed that the Einstein relation holds for the diffusivities. The use of the Einstein relation, though universally accepted, is not totally correct since it applies only strictly to equilibrium. Improvement over the Einstein relation necessitates determining the mobility and diffusivity using more exact methods of solution of the Boltzmann equation, i.e., the ensemble Monte Carlo technique. Presently, this information is not available to us and thus for simplicity, the Einstein relation is adopted as is typically done in drift-diffusion solutions.

The model's modifications made to handle heterojunctions follow the same approach as given by Sutherland and Hauser [6]. In addition, a thermionic-field emission boundary condition is used to specify the current density at heterojunction interfaces. The thermionic emission boundary condition [7], [8] is implemented in parallel with the drift-diffusion model. The actual current across the heterointerface is limited by

either the thermionic emission or diffusion current, depending upon which is smaller. In the present model, the drift-diffusion and thermionic emission current densities at the heterointerface are computed and compared. The actual current density across the heterointerface is then given by the smaller of these two quantities ensuring that the current across the junction is either thermionic emission or diffusion limited. Tunneling across the heterojunction is modeled based on the theory of Crowell and Rideout [9] and as developed in [10]. The tunneling current density is added at the heterojunction interface to give the total current density across the interface.

Equations (1)–(6) along with the proper boundary conditions can be solved for the three fundamental variables Ψ , n , and p . The Dirichlet boundary conditions of the electrostatic potential used at the Schottky contacts are

$$\psi = \psi_b + \psi_{app} - \psi_s \quad (9)$$

where Ψ_b is the built-in potential, Ψ_{app} is the applied bias voltage, and Ψ_s is the Schottky barrier height. From the thermionic emission and diffusion theory of Crowell and Sze [11] the carrier concentrations at the Schottky contacts are specified in terms of the current density passing through them as [3], [4], [12]:

$$\begin{aligned} J_n \cdot \hat{n} &= -qv_n(n - n_o) \\ J_p \cdot \hat{n} &= qv_p(p - p_o) \end{aligned} \quad (10)$$

where $v_{n,p}$ are the electron/hole thermionic recombination velocities, \hat{n} is the unit normal vector, and n_o and p_o are the equilibrium electron and hole carrier recombination at the Schottky contacts. Specifically, n_o and p_o are given as

$$n_o = N_c e^{(-q\psi/k_B T)} \quad (11)$$

$$p_o = N_v e^{(q(\psi - E_{gap})/k_B T)} \quad (12)$$

where N_c and N_v are the electron and hole effective density of states, and E_{gap} is the energy gap. At any interface, Gauss's law can be applied to relate the normal component of the electric flux density to the interface charge as

$$\epsilon_{mat1} \cdot \frac{\partial \psi}{\partial n} \Big|_{mat1} - \epsilon_{mat2} \cdot \frac{\partial \psi}{\partial n} \Big|_{mat2} = Q_{INT} \quad (13)$$

where Q_{int} is the interface charge. The free surface carrier concentrations are determined from conditions on the normal current [4], [12]:

$$J_n \cdot \hat{n} = q \cdot R^{surf} \quad (14)$$

$$J_p \cdot \hat{n} = q \cdot R^{surf} \quad (15)$$

where R^{surf} is the surface recombination rate. Equations (12)–(14) reduce to the usual Neumann boundary conditions when the interface charge and the surface recombination rates vanish.

The box integration approach of the finite difference method [4] is used to discretize (1)–(6). These equations are then solved with their appropriate boundary conditions on a nonuniform, two-dimensional mesh consisting generally of 75×33 points. The Scharfetter–Gummel technique [4], [13] is applied in the formulation of the discretization equations. The resulting

system of equations is linearized using Newton's method. Due to storage and computation time limitations the Gaussian elimination method is deemed inappropriate for the solution of the linearized system. Therefore, the system of equations is solved using an iterative approach known as the bi-conjugate gradient squared (BICGS) method [14], [15]. Only nonzero elements in the matrix are stored using this approach greatly reducing the data storage requirements of the code. After the program reaches convergence, the potential and carrier concentration profiles are readily available, from which most of the macroscopic variables of interest can be calculated. The current densities, J_n and J_p , are calculated from (4) and (5) using Scharfetter–Gummel's exponential scheme [13] for the carrier concentrations.

III. CALIBRATION OF THE MODEL

To demonstrate the accuracy of the model described above it may seem at first that direct comparison to experimental measurements would be best. However, this is not an easy task since information about the experimental setup is actually needed to accurately compare the theoretical results to experimental measurements. Different authors [5], [16] have attempted comparing the calculated response of GaAs MSM photodetectors from the drift diffusion model to experimental measurements. Landheer *et al.* [16] found that an equivalent circuit model for the experimental setup is needed to accurately compare the theoretical results to experimental measurements. The output current predicted by the circuit model of Landheer *et al.* [16] did not lead to very accurate agreement with the experimental measurements. This discrepancy is apparently due to the simplistic treatment of the external circuit and the lack of complete knowledge of the experimental circuit parameters. Alternatively, Sano [5] proposed an analytical model for GaAs MSM photodetectors based on the solution of the drift diffusion equations (1)–(6). The analytical model used is an equivalent RC circuit model in which the resistance, R , and the capacitance, C , are calculated based on the electric field and carrier concentrations obtained from the drift diffusion solution. The equivalent circuit model was implemented in a SPICE-like circuit simulator and the transient responses from the circuit simulation were compared with measured responses for an MSM photodetector. Though Sano [5] achieved good agreement between the circuit simulator and the experimental measurements, this is accomplished by numerically adjusting the circuit parameters in the model to ensure agreement with the experiment. Therefore, direct comparison of the drift diffusion results to experiment is hampered by the insufficient information about the experimental setup used in the measurements. In the absence of information about the experimental setup we choose to compare our model instead to other existing numerical models.

We have compared the results obtained using the present model against several existing device simulators and against a 1-D analytical solution for a simple $1 \mu\text{m}$ GaAs p - n junction diode. The donor and acceptor concentrations chosen for this test device are $N_D = N_A = 10^{17} \text{ cm}^{-3}$ and the carrier mobilities are $\mu_n = 7000 \text{ cm}^2/\text{V}\cdot\text{s}$ and $\mu_p = 300 \text{ cm}^2/\text{V}\cdot\text{s}$.

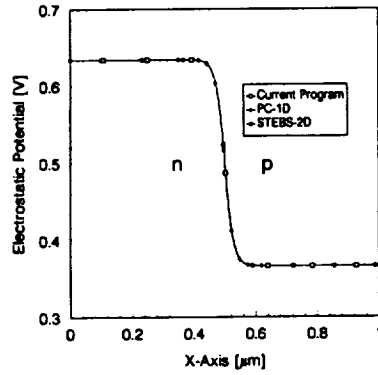


Fig. 2. Comparison of the electrostatic potentials obtained from three different numerical solutions for a GaAs pn junction diode forward biased at 1 V: current program (circles), PC-1D (diamonds), and STEBS-2D (squares).

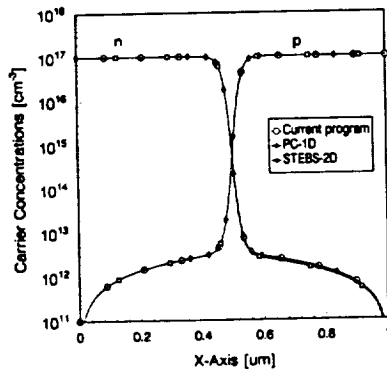


Fig. 3. Comparison of the electron and hole concentrations for a GaAs pn junction diode forward biased at 1 V calculated from the current program (circles), PC-1D (diamonds), and STEBS-2D (squares).

Fig. 2 shows the calculated electrostatic potential Ψ obtained from the present model and from two other simulators, PC-1D [17] and STEBS-2D [18]. PC-1D is a commercial, one-dimensional, drift-diffusion model released by Sandia National Laboratory and STEBS-2D is a full hydrodynamic model developed at the Georgia Institute of Technology. The carrier concentrations obtained from the three models are also presented in Fig. 3. As can be seen from inspection of Figs. 2 and 3, the present program shows precise agreement with the other two models under comparable conditions.

The electrostatic potential within the p-n junction diode calculated using the present model can also be compared to a simple one-dimensional analytical solution of Poisson's equation using the depletion approximation. Inspection of Fig. 4 shows good agreement between the numerical model and the analytical model except near the edge of the depletion region where the depletion approximation fails to properly account for the tails in the carrier distributions.

IV. APPLICATION TO MSM PHOTODETECTORS

The present numerical model is applied to study the response of GaAs based metal-semiconductor-metal (MSM) photodetectors. The general device structure considered is sketched in Fig. 1 above. The barrier height assumed for the Schottky contacts in the structure is 0.7 V and the GaAs material is taken as semi-insulating with a background doping concentration of $N_D = 10^{15} \text{ cm}^{-3}$. The interdigitated

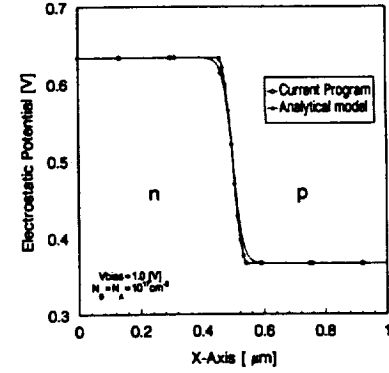


Fig. 4. Comparison of the electrostatic potentials for a GaAs pn junction diode forward-biased at 1 V calculated from the present model (circles) and a one-dimensional analytical solution (diamonds) using the depletion approximation.

TABLE I
MATERIAL PARAMETERS USED

Parameter	Units	GaAs	Ref.	AlGaAs	Ref
Electron Mass (m_n^*)	—	0.061	[18]	0.088	[18]
Hole Mass (m_p^*)	—	0.48	[18]	0.56	[18]
Dielectric Constant (ϵ)	—	13.1	[18]	12.4	[18]
Energy Gap (E_g)	(eV)	1.42	[18]	1.74	[18]
Electron Affinity (χ)	(eV)	4.06	[19]	3.80	[19]
Electron Mobility (μ_n)	(cm^2/Vs)	7000	[19]	2500	[19]
Hole Mobility (μ_p)	(cm^2/Vs)	300	[19]	150	[19]
Electron Lifetime (τ_n)	(s)	10^{-7}	[20]	10^{-8}	[19]
Hole Lifetime (τ_p)	(s)	10^{-7}	[20]	10^{-8}	[19]
Radiative coef. (B)	(cm^3/s)	2.04×10^{-10}	[21]	2.04×10^{-10}	^a
Electron Auger Coef (C_n)	(cm^6/s)	1.6×10^{-29}	[21]	1.6×10^{-29}	^a
Hole Auger Coef. (C_p)	(cm^6/s)	4.64×10^{-29}	[21]	4.64×10^{-29}	^a
Absorption Coef. (α_{abs})	(cm^{-1})	10^4	[22]	0	^b

^aDue to lack of information for AlGaAs, GaAs values are used.

^bNo absorption takes place for the AlGaAs at $\lambda = 0.84 \mu\text{m}$.

TABLE II
INPUT SIGNAL PARAMETERS

Peak Power (P_p)	0.68 mW/cm ²
Wavelength (λ)	840 nm
Beam Diameter (D)	60 μm
Beam's Peak Position (X_o)	2.5 μm
Signal Peak (t_o)	10 ps
Full Width Half Max. (FWHM)	5 ps

metallic finger widths and spacings are 1 μm and 3 μm , respectively. The material parameters used in the calculation, i.e., the zero field carrier mobilities, lifetimes, etc. are compiled in Table I. These parameters are compiled from [19]–[23]. The parameters used to characterize the optical input signal are collected in Table II. In these calculations the boundary condition at the bottom surface and along the sides of the device is assumed to be floating.

The response of the GaAs MSM device as sketched in Fig. 1 is compared to a double heterostructure device. The double heterostructure device consists of an AlGaAs barrier layer of thickness d_2 , sandwiched between two different GaAs layers, of thicknesses d_1 and d_3 , as shown in Fig. 5. The Al concentration within the AlGaAs layer is assumed to be

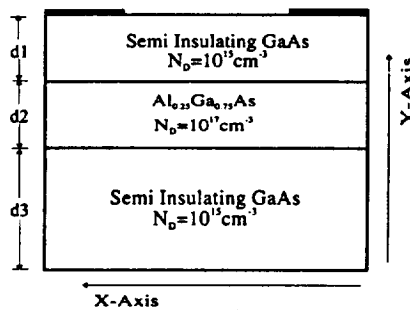


Fig. 5. Schematic diagram of the double-barrier heterostructure MSM showing the sandwiched AlGaAs layer. The solid shaded regions on the top of the figure represent the Schottky contacts.

held fixed at 25%. Different double heterostructure devices are examined with variable depths of the AlGaAs barrier layer. As a result, the active, photo-absorption layer thickness d_1 , obviously also varies in these devices.

As described above, the primary function of the buried heterostructure layers is to block the collection of the carriers photogenerated deep within the GaAs substrate by preventing them from diffusing back into the active layer and towards the contacts. In addition, at low applied bias, the barrier between the active and the AlGaAs layers acts to confine the photogenerated carriers within the active region. At higher applied bias, the heterostructure blockage of the photogenerated carriers becomes less effective due to the much greater band bending present in the device. Subsequently, the location of the AlGaAs layer greatly effects the charge collection attributes of the device depending upon the field distribution and the applied bias. In the structures examined here, the background doping is n-type, implying that the primary photogenerated carriers collected are holes.

The calculated impulse response at different active layer thicknesses, along with the corresponding rise and fall times of the signal are displayed in Fig. 6. Five different structures are examined, four double heterostructure devices and one GaAs bulk device 6 μm in thickness. The double heterostructure devices consist of a top GaAs layer, d_1 , ranging in thickness from 0.5 to 4 μm , a 1 μm AlGaAs layer, d_2 , followed by a GaAs epilayer, d_3 , ranging in thickness from 4.5 to 1 μm . Notice that the total width of all three layers combined remains constant at 6 μm . The rise and fall times are defined as the time it takes the output signal to go from 10% to 90% and from 90% to 10% of its maximum value, respectively. The voltage applied to the device is -5 V. As can be seen from Fig. 6, the fastest response occurs for the device configuration with a 0.5 μm active layer thickness. This is obvious from both the curve corresponding to the 0.5 μm device as well as from its corresponding fall time. However, the maximum output signal magnitude for the 0.5 μm device is significantly less than for the other cases. This is as expected, since the response speed is achieved at the expense of lower output signal magnitude because the slower carriers, those generated deep within the device, are blocked from being collected by the heterojunction barrier. As the active layer thickness d_1 increases to 1 μm , more carriers are generated within the top, active GaAs region. As a result, a greater number of

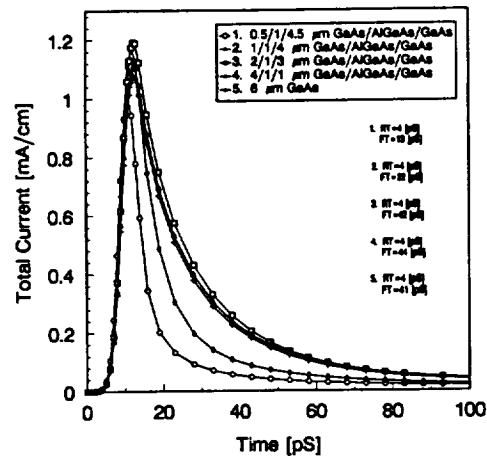


Fig. 6. The calculated impulse response at different active layer thicknesses for the double-barrier MSM heterostructure at an applied bias of -5 V. The doping levels are 10^{15} cm^{-3} for the GaAs layers and 10^{17} cm^{-3} for the AlGaAs layer. Curve 5 is for the case without the AlGaAs layer. The numbers in the legend represent the layer thickness in microns of the GaAs/AlGaAs/GaAs layers as shown in Fig. 5. Also included are the rise and fall times for each of the 5 curves.

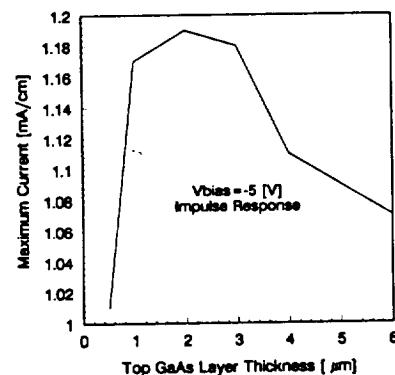


Fig. 7. The maximum output current signal as a function of active layer thickness for the double-barrier MSM heterostructure.

photogenerated carriers are collected producing a higher output signal current. Though the $d_1 = 1 \mu\text{m}$ device does not show as rapid a collection of the photogenerated carriers as the $d_1 = 0.5 \mu\text{m}$ device, the field is sufficiently strong and the carriers are still relatively close to the collecting contacts that a reasonably high speed of response is retained; a fall time of 22 ps is achieved as compared to 13 ps for the 0.5 μm device. If the active layer thickness is increased further to 2 μm , the same trend is observed; more carriers are collected from the bottom of the active layer resulting in a longer fall time, ~ 42 ps, and slower speed of response. Interestingly, the maximum output signal current ultimately decreases with increasing d_1 . The maximum signal current is plotted as a function of active layer thickness in Fig. 7. As can be seen from Fig. 7, the output signal current reaches a maximum for an active layer thickness of $d_1 = 2 \mu\text{m}$. The maximum signal current is significantly less for a device with $d_1 = 4 \mu\text{m}$ than with $d_1 = 2 \mu\text{m}$, though the fall times are comparable. Clearly, at an applied bias of -5 V, there exists an optimal thickness, in terms of speed of response and collection efficiency, for layer d_1 .

In an attempt to understand the origin of the peak in the output current signal versus active layer thickness curve, we

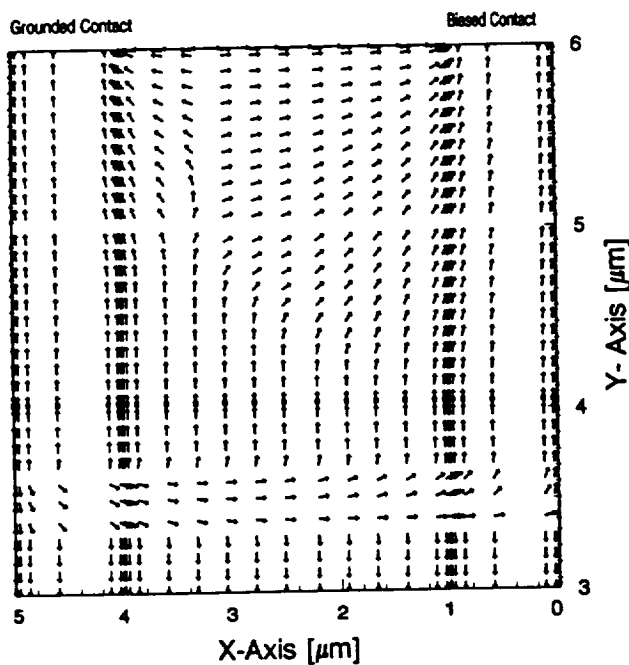


Fig. 8. Electric field vector plot at the time-step corresponding to the peak current showing the direction of the electric field at each mesh point within the two-dimensional simulation grid used for the $2.0\ \mu\text{m}$ active layer thickness device. The AlGaAs layer is located $2.0\ \mu\text{m}$ from the top and is $1.0\ \mu\text{m}$ in thickness. Only the top $3\ \mu\text{m}$ of the device is shown in the figure. The right top contact of the device is biased at $-5\ \text{V}$ with respect to the left top contact. The horizontal channeling of the electric field at the middle of the heterobarrier is due to the accumulation of electrons within the AlGaAs trapped by the potential band bending at either interface. Notice that the lateral component of the electric field vanishes within the active region above the heterobarrier.

have examined the electric field profiles within each device. Field vector plots showing the direction of the electric field at each mesh point within the device are shown in Figs. 8 and 9 for the 2.0 and $4.0\ \mu\text{m}$ active layer thickness devices, respectively. Comparison of the two figures clearly shows that the lateral field component essentially vanishes near the heterostructure in the $2.0\ \mu\text{m}$ width device, while a significant lateral component persists within the $4.0\ \mu\text{m}$ width device. The electric field points vertically away from the heterostructure acting to accelerate the photogenerated holes towards the collecting contacts in the $2.0\ \mu\text{m}$ device. Alternatively, in the $4.0\ \mu\text{m}$ device, the lateral component of the electric field acts to accelerate the holes to some extent laterally. As a result, the hole trajectory for collection is longer, effectively delaying their collection at the contacts. Since the holes spend more time in this region their chances of suffering a recombination event increase, leading to a reduction in the collected current. Subsequently, this could result in an increased collected current within the $2.0\ \mu\text{m}$ width device than in the $4.0\ \mu\text{m}$ width structure. A more detailed analysis of this problem using a more sophisticated, hydrodynamic model will be made in the future to further test this hypothesis.

The effect of the heterojunction on the response to a square wave input is considered next. The same model and approach as described above is used but with a square wave input signal applied to the device. The output current response to a square wave input with repetition rate of $1\ \text{GHz}$ and $0.5\ \text{ns}$ duration time is shown in Fig. 10 for different GaAs active

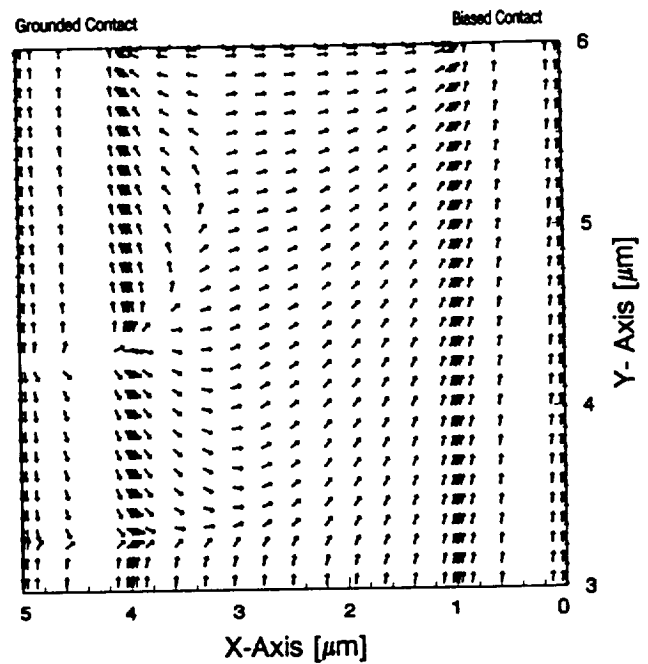


Fig. 9. Electric field vector plot at the time-step corresponding to the peak current showing the direction of the electric field at each mesh point within the two-dimensional simulation grid used for the $4.0\ \mu\text{m}$ active layer thickness device. The AlGaAs layer is located at $4.0\ \mu\text{m}$ from the top and is $1.0\ \mu\text{m}$ in thickness. Only the top $3\ \mu\text{m}$ of the device is shown in the figure. The right top contact of the device is biased at $-5\ \text{V}$ with respect to the left top contact. Notice that the electric field "swirls" around deeper within the active region of the device. The presence of the lateral component of the electric field acts to increase the trajectory of the carriers leading to a longer time between generation and collection with the subsequent increase of recombination losses.

layer thicknesses. The applied voltage is again $-5\ \text{V}$. It is clear from this figure that the output signal amplitude peaks at an active layer thickness of $2\ \mu\text{m}$ and then decreases as the active layer gets thicker. The explanation for this is similar to that described above for the impulse response. The rise and fall times however are not quite consistent with the impulse response results. This is due primarily to the definition of these quantities. Although the $0.5\ \mu\text{m}$ curve in Fig. 10 seems to decay faster than any of the other cases, the fall time, for example, is not the shortest. This is due to the fact that the 90–10% measure of the fall time depends on the magnitude of the signal peak itself. Since the signal peak is very much smaller in the $0.5\ \mu\text{m}$ device, its decay to 10% takes longer than that for a much higher signal peak, though the signal in the $0.5\ \mu\text{m}$ device has decayed quicker below some threshold level. Subsequently, the definition of the rise and fall times typically given, 10–90% and 90–10% respectively, are of questionable value when evaluating the performance of an MSM detector subject to a square wave input. However, it is clear that the presence and the location of the AlGaAs layer is critical to the performance of these photodetectors and that the optimal location of the heterojunction barriers is a function of the applied bias. This is clearly demonstrated in Fig. 11 where the applied voltage is increased to $-10\ \text{V}$. We note that in this case the peak in the collected current shifts to the $3\ \mu\text{m}$ active layer thickness device. The shift in the peak collected current occurs since the field depletes deeper into the semiconductor

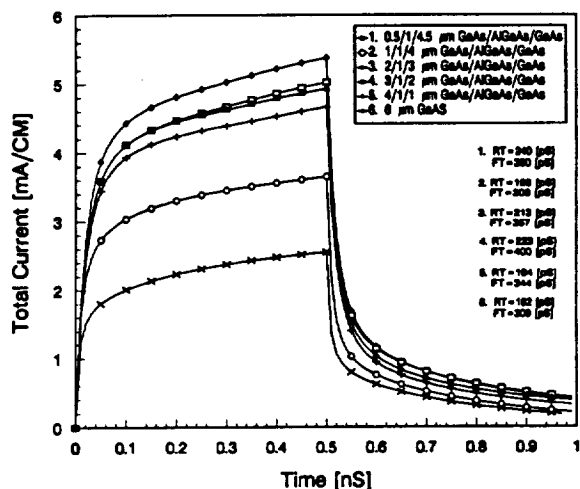


Fig. 10. The calculated output current response of the double-barrier MSM heterostructure to a square-wave input at different active layer thicknesses and under an applied voltage of -5 V. Curve 6 is for the case without the AlGaAs layer. The numbers in the legend represent the layer thickness in microns for the GaAs/AlGaAs/GaAs layers as shown in Fig. 5. Also included are the rise and fall times in picoseconds for each of the curves.

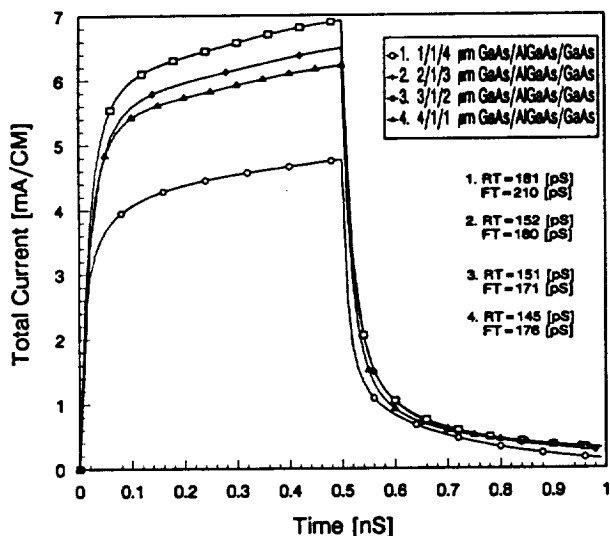


Fig. 11. The calculated output current response of the double-barrier MSM heterostructure to a square-wave input at different active layer thicknesses and under an applied bias of -10 V. Curve 6 is for the case without the AlGaAs layer. The numbers in the legend represent the layer thickness in microns of the GaAs/AlGaAs/GaAs layers as shown in Fig. 5. Also included are the rise and fall times in picoseconds for each of the curves.

material. Subsequently, more carriers are swept out by the relatively high-field producing a greater collected current at the contacts.

Due to computational limitations, it is presently possible to simulate the entire substrate. Subsequently, only a small portion of the substrate can be simulated. In this case, the bottom surface can be treated either as floating or as an ohmic contact. The use of an ohmic boundary condition along the bottom surface leads to qualitatively similar results as those with the floating surface condition but with some quantitative difference. The appropriateness of these two boundary conditions needs to be further investigated, which will be reported in a future work.

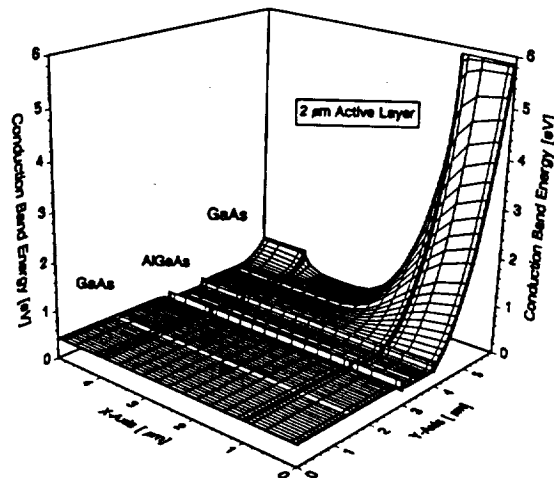


Fig. 12. Two-dimensional conduction band diagram of the double-barrier MSM heterostructure for the $2 \mu\text{m}$ active layer thickness device at an applied bias of -5 V.

Finally, we examine the dark current for the device sketched in Fig. 1 assuming a finger length of $100 \mu\text{m}$, typical for most MSM structures. The dark current is found to be on the order of 0.6 pA . Moreover, it is also found that it is practically insensitive to the applied bias in the range from zero to -20 V. This is because, at a $3 \mu\text{m}$ finger spacing, and at the applied biases considered here, the built-in potential barrier at the grounded Schottky contact persists as shown in the two-dimensional conduction band plot for the $2.0 \mu\text{m}$ active layer thickness device in Fig. 12. However, if the voltage is increased or if the finger spacing is reduced, the applied voltage at the biased contact punches through to the grounded contact reducing the built-in potential barrier. As a result, an increased electron dark current develops. It is expected then that for a smaller finger spacing ($\sim 1 \mu\text{m}$) and a doping level of 10^{15} cm^{-3} that the dark current will increase with increasing applied voltage [24], [25]. In that case, a top AlGaAs layer [26] can play an important role in limiting the dark current of the device.

V. CONCLUSIONS

We have theoretically investigated the effect of including a double heterostructure barrier on the response of MSM interdigitated photodetectors using a two-dimensional drift-diffusion model with a thermionic-field emission boundary condition. To establish the validity of the model we have compared it to existing models and to a one-dimensional analytical solution. Excellent agreement with these models is obtained. The drift-diffusion model is then used to study the effect of the presence of a sandwiched AlGaAs layer on the response, as measured by the responsivity and speed, of MSM photodetectors. The AlGaAs layer introduces a double heterostructure barrier whose location in the structure greatly affects the movement of the carriers. We have found that there exists an optimal location for these heterojunction barriers for which both high speed and high responsivity can be achieved for an input optical signal of 840 nm wavelength. The placement of the AlGaAs barrier depends on the applications in which the MSM's are to be used. In high speed applications,

and under low bias conditions, the AlGaAs layer should be placed near the top surface of the device. To obtain a fast response under low light illumination levels, the active layer thickness and the applied bias should be increased. However, operation at high bias can increase the dark current levels of the device if the electrode spacing is small. Subsequently, there exist several tradeoffs in the design of an interdigitated MSM photodetector. Use of a simulator such as the one described here, is essential for optimizing a structure for a particular application.

ACKNOWLEDGMENT

The authors would like to thank Steven H. Janzou for help in developing the mesh refinement for the drift-diffusion calculation and Dr. Tom Oppe at Florida State University for setting up the NSPCG solver routines. The authors would also like to thank Drs. Mahmoud Bahrami and Paul Vella of the Digital Equipment Corporation for many useful discussions during the course of this work.

REFERENCES

- [1] S. Y. Chou and M. Y. Liu, "Nanoscale tera-hertz metal-semiconductor-metal photodetectors," *IEEE J. Quant. Electron.*, vol. 28, pp. 2358-2368, Oct. 1992.
- [2] L. Figueroa and C. W. Slayman, "A novel heterostructure interdigitated photodetector (HIP) with picosecond optical response," *IEEE Electron. Dev. Lett.*, vol. EDL-2, pp. 208-210, Aug. 1981.
- [3] G. Baccarani, M. Rudan, R. Guerrieri, and P. Ciampolini, "Physical model for numerical device simulation," in *Process and Device Modeling*, W. L. Engl, Ed. Amsterdam: North Holland, 1986, pp. 107-158.
- [4] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Wien, New York: Springer-Verlag, 1984.
- [5] E. Sano, "A device for metal-semiconductor-metal photodetectors and its applications to optoelectronic integrated circuit simulation," *IEEE Trans. Electron Devices*, vol. 37, pp. 1964-1968, Sept. 1990.
- [6] J. E. Sutherland and J. R. Hauser, "A computer analysis of heterojunction and graded composition solar cells," *IEEE Trans. Electron Devices*, vol. ED-24, pp. 363-372, Apr. 1977.
- [7] C. M. Wu and E. S. Yang, "Carrier transport across heterojunction interfaces," *Solid State Electron.*, vol. 22, pp. 241-248, 1979.
- [8] K. Horio and H. Yanai, "Numerical modeling of heterojunctions including the thermionic emission mechanism at the heterojunction interface," *IEEE Trans. Electron Devices*, vol. 37, pp. 1093-98, Apr. 1990.
- [9] C. R. Crowell and V. L. Rideout, "Normalized thermionic-field emission in metal-semiconductor (Schottky) barriers," *Solid-State Electron.*, vol. 12, pp. 89-105, 1969.
- [10] K. Yang, J. R. East, and G. Haddad, "Numerical modeling of abrupt heterojunctions using a thermionic-field emission boundary condition," *Solid State Electron.*, vol. 36, pp. 321-330, 1993.
- [11] C. R. Crowell and S. M. Sze, "Current transport in metal-semiconductor barriers," *Solid State Electron.*, vol. 9, pp. 1035-1048, 1966.
- [12] W. L. Engl, H. K. Dirks, and B. Meinerzhagen, "Device modeling," *Proc. IEEE*, vol. 71, pp. 10-33, Jan. 1983.
- [13] D. L. Scharfetter and H. K. Gummel, "Large-signal analysis of a silicon read diode oscillator," *IEEE Tran. Electron Devices*, vol. ED-16, pp. 64-77, Jan. 1969.
- [14] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore: Johns Hopkins Univ. Press, pp. 362-379, 1983.
- [15] T. C. Oppe, W. D. Joubert, and D. R. Kincaid, "An overview of NSPCG: a nonsymmetric preconditioned conjugate gradient package," *Computer Physics Commun.*, vol. 53, pp. 283-293, 1989.
- [16] D. Landheer, Z.-M. Li, S. P. McAlister, and D. A. Aruliah, "Modeling of ultrafast metal-semiconductor-metal photodetectors," *Can. J. Phys.*, vol. 69, 1991.
- [17] P. A. Basore, D. T. Rover, and A. W. Smith, "PC-1D version 2; enhanced numerical simulation," *Proc. 20th IEEE Photovoltaic Spec. Conf.*, IEEE, New York, 1988, p. 389.
- [18] A. W. Smith, "Light confinement and hydrodynamic modelling of semiconductor structures by volumetric methods," Ph.D. dissertation, Georgia Inst. of Technol., Atlanta, GA, 1992.
- [19] H. C. Casey, Jr. and M. B. Panish, *Heterostructure Lasers*. New York: Academic, 1978.
- [20] K. Horio and H. Yanai, "Numerical modeling of heterojunctions including the thermionic emission mechanism at the heterojunction interface," *IEEE Trans. Electron Devices*, vol. 37, pp. 1093-1098, Apr. 1990.
- [21] M. Bahrami, private communication, Digital Equipment Corp.
- [22] S. A. Ringel, "Growth and process induced defects and recombination mechanisms in AlGaAs/GaAs and CdZnTe/CdS photovoltaic device structures," Ph.D. dissertation, Georgia Inst. of Technol., Atlanta, GA, 1992.
- [23] S. M. Sze, *Physics of Semiconductor Devices*. New York: Wiley, 1981, p. 750.
- [24] M. Ito and O. Wada, "Low dark current GaAs metal-semiconductor-metal (MSM) photodiodes using WSi_x contacts," *IEEE J. Quant. Electron.*, vol. QE-22, pp. 1073-1077, 1986.
- [25] D. S. Malhi, J. M. Xu, F. Hegmann, B. Takasaki, and R. Surridge, "Effects of material modification on dark current of GaAs MSM photodetectors," *6th Conf. on Semi-insulating III-V Materials*, 1990, pp. 451-456, Toronto, Canada.
- [26] A. Aboudou, J. P. Vilcot, D. Decoster, A. Chenoufi, E. Delhaye, P. Boissenot, C. Varin, F. Deschamps, and I. Lecuru, "Ultra-low dark current GaAlAs/GaAs MSM photodetector," *Electronics Lett.*, vol. 27, no. 10, pp. 793-95, May 1991.



Ali F. Salem (S'90) was born in Marabon, Lebanon, on September 2, 1967. He received the B.S. degree from Boston University, Boston, MA, in 1989 and the M.S. degree from the Georgia Institute of Technology, Atlanta, in 1990, both in electrical engineering. He is currently pursuing his Ph.D. degree in electrical engineering at the Georgia Institute of Technology. His dissertation work involves modeling and simulation of carrier transport in semiconductor devices and its applications.

Arlynn W. Smith received the B.S. degree in ceramic engineering from Alfred University in 1984, the M.S. degree from Georgia Institute of Technology in 1987 and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, in 1992. From 1992 to 1993 he held a post-doctoral fellowship at the Georgia Tech Research Institute, where he developed a three dimensional simulation code for modelling the chemical vapor infiltration of ceramic matrix composites. Currently, he is performing a post-doctoral research program with the microelectronics research center at the Georgia Institute of Technology developing a hydrodynamic simulation code of heterojunction acoustic charge transport devices, charge transport devices, and avalanche photodiodes.



Kevin Brennan (S'84-M'84-SM'90) received the B.S. degree in physics from the Massachusetts Institute of Technology, Cambridge, MA, in 1978, and the M.S. degrees in physics in 1980, and the Ph.D. degree in electrical engineering from the University of Illinois, Urbana-Champaign, IL, in 1984.

He is currently a Professor of Electrical and Computer Engineering at the Georgia Institute of Technology, Atlanta, GA. His current research interests include the theory and modeling of avalanche

devices and confined state photomultipliers, the physics of impact ionization, nonlinear transport effects in semiconductors and heterostructures, plasmons in semiconductors, and the theory and modeling of acoustic charge transport devices.

Dr. Brennan was awarded a Presidential Young Investigator Award through the National Science Foundation in 1988.

Theoretical Study of the Response of InGaAs Metal-Semiconductor-Metal Photodetectors

Ali F. Salem, *Student Member, IEEE*, and Kevin F. Brennan, *Senior Member, IEEE*

Abstract— We present a theoretical investigation of the response of metal-semiconductor-metal (MSM) photodetectors made of InGaAs lattice-matched to InP using a two-dimensional drift-diffusion model with a thermionic-field emission boundary condition for the heterojunctions. The effect of including a top InAlAs layer to increase the effective barrier height of the metal fingers on the InGaAs active layer is thoroughly examined and found to limit the collection of the photocurrent signal due to the electron and hole barriers that it forms with InGaAs. Due to the thickness and height of the InAlAs barrier layer in existing designs, the tunneling current obtained from the model is found to be negligibly small to significantly affect the output signal current. In an attempt to obtain a better response, different design structures including one where a quasi-Schottky contact is utilized are studied and their speed of response, breakdown voltage, and dark current are compared to that of the usual InGaAs device.

I. INTRODUCTION

PLANAR metal-semiconductor-metal (MSM) photodetectors are becoming an important component in integrated circuit technologies owing to their ease of fabrication, compatibility with existing field-effect transistors (FET's), and high responsivity-bandwidth performance [1]. High-performance GaAs MSM's have been thoroughly investigated [2]–[4] and found to have an excellent performance in the $0.8\ \mu\text{m}$ wavelength region. For wavelengths in the low-loss window of silica fiber-optic communication lines (1.3 – $1.5\ \mu\text{m}$), the lower bandgap material, InGaAs, lattice-matched to InP, has been under intensive examination [5]–[8]. Unfortunately, the low barrier height on undoped InGaAs ($\sim 0.2\ \text{V}$) yields unacceptably large dark currents in photodetectors made with this material leading to poor overall device performance.

In order to reduce the device dark current, several schemes have been suggested to enhance the effective barrier height of InGaAs MSM's by growing a thin layer of a high-barrier height material on top of the active InGaAs layer [5]–[8]. One of these schemes incorporates a thin, lattice matched InAlAs top layer that has a barrier height of $\sim 0.65\ \text{eV}$. A schematic diagram of this device structure is shown in Fig. 1. The effective barrier height obtained from the InAlAs layer limits the dark current to acceptable levels but it introduces, in

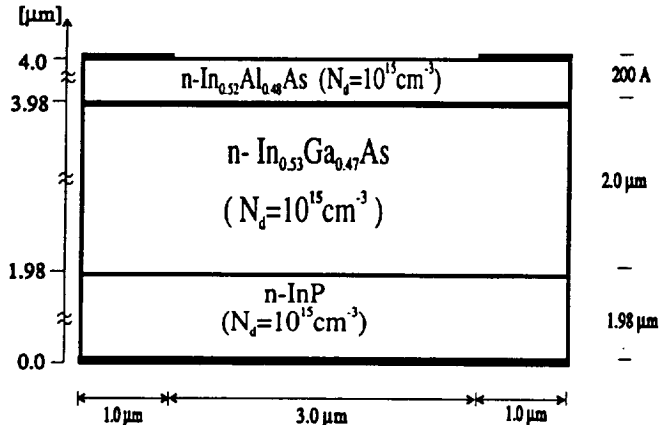


Fig. 1. Cross-sectional view of the simulated InGaAs based MSM with an InAlAs top layer to enhance the effective Schottky contact barrier height of InGaAs. The top dark-shaded areas are Schottky contacts and the bottom dark-shaded area is an ohmic contact.

turn, large barriers for electrons and holes [5] thereby limiting the collection of the photogenerated carriers and hence the performance of the device.

The performance of the InAlAs–InGaAs–InP device sketched in Fig. 1 under both dark and illumination conditions has been studied experimentally by several authors [5]–[7]. In order to illustrate the physical workings within the device, Soole and Schumacher [5] used a simple one-dimensional thermionic emission model to probe the transport physics in this structure. They concluded that the InAlAs raises the effective barrier height to values close to the barrier height on InAlAs and that the experimentally observed soft breakdown is due to tunneling across the barrier under the Schottky contacts. They have also indicated that a two-dimensional analysis with both thermionic emission and tunneling being incorporated is necessary to understand the nature of the transport in these devices.

It has also been reported [9] that the barrier height on n-InGaAs could be raised by growing a thin, oppositely-doped p^+ assist layer between the Schottky metal contacts and the n-InGaAs layer to form what is called “quasi-Schottky” contacts. Fig. 2 shows a schematic side-view of such a device. In Fig. 2, the n^+ -InP layer acts as a buffer and the presence of the heavily doped n^+ -InGaAs layer is introduced to prevent the device from acting as a rectifier at the InGaAs–InP interface. Averin *et al.* have analyzed and experimentally confirmed [9] the formation of the quasi-Schottky contacts of Fig. 2. They have found that for a properly designed p^+ layer the effective barrier height on n-InGaAs can be increased from $0.2\ \text{eV}$ to $\sim 0.54\ \text{eV}$.

Manuscript received September 21, 1994; revised November 24, 1994. This work was supported in part by the National Science Foundation through a Presidential Young Investigator Award made to K. Brennan, by the Digital Equipment Corp. through contract E21-H36, and by ARPA through contract to NASA, NAGW-2753.

The authors are with the School of Electrical and Computer Engineering and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

IEEE Log Number 9409833.

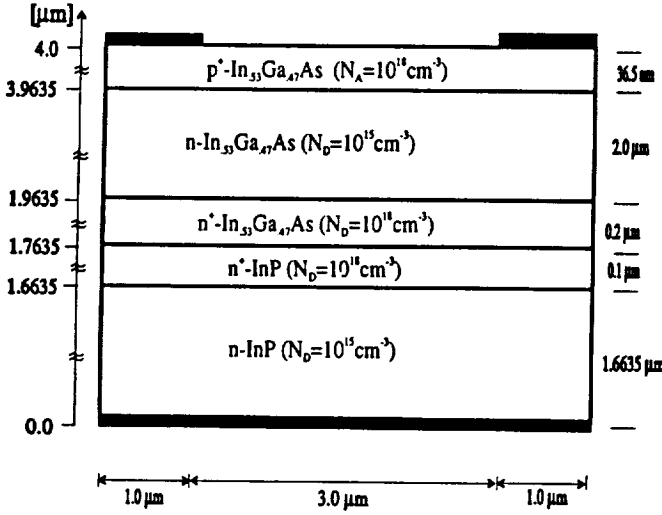


Fig. 2. Cross-sectional view of the simulated p^+ assist, quasi-Schottky device. The top dark-shaded areas are Schottky contacts and the bottom dark-shaded area is an ohmic contact.

Recently, Lee *et al.* [10] reported that Schottky barrier heights on n-InGaAs can be improved by metal deposition on substrates that are cooled to a temperature of 77 K. They have found that the barrier height on n-InGaAs can be increased from 0.2 eV to ~ 0.6 eV with the use of Ag metal. Fig. 3 shows a schematic diagram of an InGaAs interdigitated MSM photodetector built on an InP substrate. The Schottky contacts in Fig. 3 are deposited on the low temperature n-InGaAs layer to improve the Schottky barrier height. It can be seen from the device of Fig. 3 that fewer processing steps are required than the devices of Figs. 1 and 2.

In this paper, we present a theoretical investigation of the issues involved in designing an InGaAs-based interdigitated MSM photodetector. The different approaches shown in Figs. 1–3 to enhance the Schottky barrier heights on n-InGaAs are numerically investigated and their performance is compared. The model used herein is a two-dimensional, numerical, drift-diffusion calculation that incorporates both thermionic emission and tunneling across heterobarriers. This model is used to predict the speed of response, dark current, and breakdown voltage for the devices of Figs. 1–3. The effect of the InAlAs–InGaAs heterobarrier of Fig. 1 on the detector's response is thoroughly examined using the present model and the performance of this device is compared to that of the devices of Figs. 2 and 3 where such a heterobarrier is not present. Moreover, the response of the device of Fig. 2 is compared to that of the same device with the exception of using ohmic contacts instead of the existing Schottky contacts. The technological problems as well as the advantages and disadvantages of these different design structures are also evaluated.

This paper is organized as follows. In Section II, we briefly describe the numerical model used in the simulation. Specifically, we characterize the drift-diffusion model and discuss the formulations used to treat the thermionic emission and tunneling transport. In Section III, we apply this model to study the performance of three different design structures of

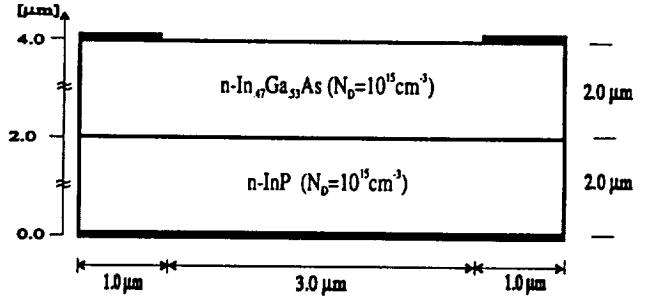


Fig. 3. Cross-sectional view of the simulated low-temperature grown Schottky contacts device where the Schottky contacts are assumed to be grown at a temperature of 77 K. The top dark-shaded areas are Schottky contacts and the bottom dark-shaded area is an ohmic contact.

InGaAs-based MSM photodetectors; the InAlAs–InGaAs–InP standard device of Fig. 1, the quasi-Schottky device of Fig. 2, and the low-temperature-formed contacts device of Fig. 3. The performance of these different designs in terms of their speed of response, dark current, and breakdown voltage is also compared in Section III. Potential band diagram plots are presented in Section III to demonstrate the effect of the different types of barriers on the response of these devices. Finally, conclusions are drawn in Section IV.

II. MODEL DESCRIPTION

To understand the behavior of photo-generated carriers in MSM's under the influence of drift and diffusion forces, Poisson's equation and the electron and hole current continuity equations are solved on a two-dimensional simulation domain for the electrostatic potential and the carrier concentrations. The width and depth of the MSM detectors simulated are restricted to dimensions on the order of microns where the drift-diffusion approximation is generally sufficient to describe the transport physics. The effects of thermionic emission and tunneling on the transport are modeled as boundary conditions across heterojunction barriers.

The full details of the model have been reported elsewhere [11]. Nevertheless, the salient features of the model are summarized here. The basic equations used in the simulator are [12]:

$$\epsilon_0 \epsilon_s \nabla^2 \psi = -q(p - n + N_D - N_A) \quad (1)$$

$$-\frac{1}{q} \nabla \cdot J_n - G_n + R_n + \frac{\partial n}{\partial t} = 0 \quad (2)$$

$$\frac{1}{q} \nabla \cdot J_p - G_p + R_p + \frac{\partial p}{\partial t} = 0 \quad (3)$$

$$J_n = q\mu_n n E_n + qD_n \nabla n \quad (4)$$

$$J_p = q\mu_p p E_p - qD_p \nabla p \quad (5)$$

where Ψ is the electrostatic potential, n and p are the electron and hole carrier concentrations, $N_{D,A}$ are the donor/acceptor concentrations, $J_{n,p}$ are the electron/hole currents, $R_{n,p}$ and $G_{n,p}$ are the electron/hole recombination and generation rates,

and $\mu_{n,p}$ and $D_{n,p}$ are electron/hole mobilities and diffusivities. The terms $R_{n,p}$ used in (2) and (3) represent the sum of Shockley-Read-Hall, radiative, and Auger recombinations for electrons and holes [12]. A standard field-dependent mobility model [12] is used for electrons while a zero-field mobility model is used for holes. $D_{n,p}$ is calculated based on the Einstein relation. The generation rate, $G_{n,p}$, is the sum of both optical generation and impact ionization generation rates for electrons and holes. The optical generation rate is assumed to be uniform over the area of the basic unit cell of the devices simulated here. The optical generation rate is assumed gaussian in time and decays exponentially with depth with the absorption coefficient, α_{Abs} .

The impact ionization generation rates for electrons and holes, $G_{n,p}^{II}$, are assumed to be proportional to the magnitude of the electron and hole currents, respectively, and are given as [12]:

$$G_n^{II} = \alpha \frac{|J_n|}{q}, \quad G_p^{II} = \beta \frac{|J_p|}{q} \quad (6)$$

where α and β are the electron and hole ionization rates coefficients defined as the number of electron-hole pairs generated in one unit length of travel. These coefficients are modeled as exponential functions of the local electric field in the direction of the current flow [12]. The parameters involved are determined from fitting the experimental data [13] using the following equation:

$$\alpha = \alpha^\infty \cdot \exp\left(-\left(\frac{F_n^{crit}}{F_n}\right)^{\gamma_n}\right) \\ \beta = \beta^\infty \cdot \exp\left(-\left(\frac{F_p^{crit}}{F_p}\right)^{\gamma_p}\right) \quad (7)$$

where α and β are the electron and hole impact ionization rates, α^∞ and β^∞ are the rates at infinite electric fields, $F_{n,p}^{crit}$ are the critical electron and hole fields, and $\gamma_{n,p}$ are fitting parameters. $F_{n,p}$, the components of the local electric field F in the direction of the electron and hole currents, are described as

$$F_n = \left| \frac{J_n \cdot F}{J_n} \right|, \quad F_p = \left| \frac{J_p \cdot F}{J_p} \right| \quad (8)$$

Thermionic emission and tunneling at the heterojunction interfaces are modeled as boundary conditions [14], [15]. First, the current crossing a heterointerface is assumed to be either thermionic emission or diffusion limited. The model independently calculates both the thermionic emission and the drift-diffusion currents at a heterojunction interface and compares them. The smaller of these two values is taken as the limiting current. This limiting current, in addition to any tunneling currents across the heterobarrier, dictates the current flow for the overall device. The thermionic-tunneling boundary condition for the electron current at a heterojunction interface can be written in terms of quantities of materials 1 and 2 that form the heterojunction as [14], [15]:

$$J_n = q \left(n_2 v_{n2} (1 + \delta) - \Gamma n_1 v_{n1} (1 + \delta) \exp\left(-\frac{\Delta E_c}{kT}\right) \right) \quad (9)$$

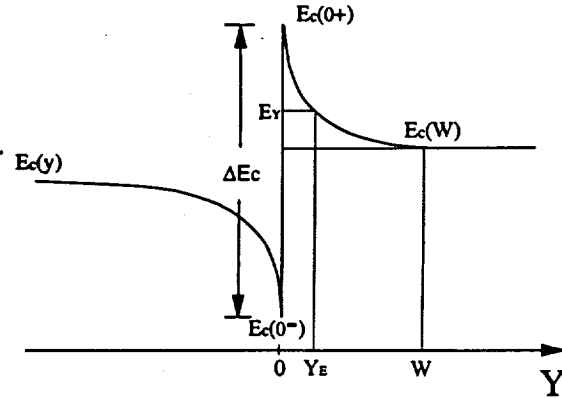


Fig. 4. Conduction band edge diagram of an abrupt heterojunction. Y_E denotes the width of the tunneling barrier at an incident energy E_y . The energy range over which tunneling takes place is determined by $E_{c,min}$ [$=\min(E_c(0^-), E_c(W))$] and $E_c(0^+)$, where W is the depletion region width on the barrier side.

where n_1 and n_2 are the electron concentrations in materials 1 and 2, v_{n1} ($=A_1^* T^2 / q N_{C1}$) and v_{n2} ($=A_2^* T^2 / q N_{C2}$) correspond to the recombination velocities of the Schottky barrier theory, ΔE_c is the conduction band edge discontinuity that is calculated based on the affinity rule, q is the elementary charge, Γ is the ratio of the masses m_2/m_1 , k is the Boltzmann constant, and T is the temperature. The parameter, δ , in (9) is a coefficient that describes the tunneling across the heterojunction interface and is given as [15]:

$$\delta = \frac{\exp\left(\frac{E_{c,max}}{kT}\right)}{kT} \int_{E_{c,min}}^{E_{c,max}} \exp\left(-\frac{E_y}{kT}\right) P(E_y) dE_y \quad (10)$$

where $E_{c,min}$ and $E_{c,max}$ determine the energy range over which tunneling is possible, and $P(E_y)$ is the tunneling probability. The tunneling probability, $P(E_y)$, is calculated based on the WKB approximation and is formulated as in [15]:

$$P(E_y) = \exp\left(-\frac{4\pi}{h} \int_0^{Y_E} [2m_n^* (E_c(y) - E_y)]^{1/2} dy\right) \quad (11)$$

where m_n^* is the effective mass of the carrier in the semiconductor and the bounds of integration, as shown in Fig. 4, determine the width of the energy barrier at an energy level E_y . The remaining parameters in (11) are indicated on Fig. 4.

The thermionic emission current, J_p , and the tunneling parameter, δ , for holes are similar to those used for electrons. Note that if δ is set equal to zero and Γ is set equal to 1 in (9) above the usual thermionic emission current boundary condition of the Schottky type is obtained.

The box integration approach of the finite difference method [12] is used to discretize equations (1)–(6). These equations are then solved with their appropriate boundary conditions on a nonuniform, two-dimensional mesh consisting generally of 65×40 points. The Scharfetter-Gummel technique [12], [17] is applied in the formulation of the discretization equations. The resulting system of equations is linearized using Newton's method. Due to storage and computation time limitations the Gaussian elimination method is deemed inappropriate and the

TABLE I

Parameter	Units	InGaAs	Ref.	InAlAs	Ref.	InP	Ref.
Electron Mass (m_e)	-	0.0463	[21]	0.084	[21]	0.08	[25]
Heavy Hole Mass (m_{hh})	-	0.61	[21]	0.677	[21]	0.85	[25]
Light Hole Mass (m_{lh})	-	0.0588	[21]	0.086	[21]	0.089	[25]
Dielectric Constant (ϵ)	-	13.85	[21]	12.42	[21]	12.61	[26]
Energy Gap (E_g)	(eV)	0.77	[21]	1.49	[21]	1.34	[26]
Electron Affinity (χ)	(eV)	4.7	[26]*	4.2	[26]*	4.51	[26]
Electron Mobility (μ_n)	(cm ² /Vs)	12000	[22]	800	[22]	5400	[26]
Hole Mobility (μ_p)	(cm ² /Vs)	300	[22]	25	[22]	300	[26]
Electron Capture Cross Section (σ_n)	(cm ²)	3×10^{-17}	[23]	3×10^{-17}	[23] ^g	7×10^{-16}	[23]
Hole Capture Cross Section (σ_p)	(cm ²)	3×10^{-17}	[23]*	3×10^{-17}	[23] ^g	7×10^{-16}	[23]*
Trap Concentration (N_t)	(cm ⁻³)	1.5×10^{13}	[23]	1.5×10^{13}	[23] ^g	3×10^{16}	[23]
Radiative coef. (B)	(cm ³ /s)	1.2×10^{-10}	[24]	1.2×10^{-10}	[24] ^g	3.32×10^{-10}	[27]
Electron Auger Coef. (C_n)	(cm ⁶ /s)	3.2×10^{-28}	[24]	3.2×10^{-28}	[24] ^g	3.7×10^{-31}	[28]
Hole Auger Coef. (C_p)	(cm ⁶ /s)	3.2×10^{-28}	[24]*	3.2×10^{-28}	[24] ^g	8.7×10^{-30}	[28]
Absorption Coef. (α_{ph}) at $\lambda = 1.55 \mu\text{m}$	(cm ⁻¹)	8000	[20]	0.0	-	0.0	-
Electron Impact ionization coefficients:							
-(α_n^-)	(cm ⁻¹)	4.677×10^7	[13]	7.36×10^4	[13]	1.12×10^7	[13]
-(F_n^{ion})	(V-cm ⁻¹)	1.935×10^6	[13]	5.26×10^5	[13]	3.11×10^6	[13]
-(γ_n)	-	1.0	[13]	2.0	[13]	1.0	[13]
Hole Impact ionization coefficients:							
-(β_p^-)	(cm ⁻¹)	8.382×10^7	[13]	1.57×10^4	[13]	4.79×10^6	[13]
-(F_p^{ion})	(V-cm ⁻¹)	2.236×10^6	[13]	4.89×10^5	[13]	2.55×10^6	[13]
-(γ_p)	-	1.0	[13]	2.0	[13]	1.0	[13]

+ Calculated based on the InAs and GaAs binaries; * Electron values used; # InGaAs values used

linearized system is solved using an iterative approach known as the bi-conjugate gradient squared (BICGS) method [18], [19]. Only nonzero elements in the matrix are stored using this approach greatly reducing the data storage requirements of the code. After the program reaches convergence, the potential and carrier concentration profiles are readily available, from which most of the macroscopic variables of interest can be calculated. The current densities, J_n and J_p , are calculated from Equations 4 and 5 using Scharfetter-Gummel's exponential scheme [17] for the carrier concentrations.

III. NUMERICAL RESULTS

The numerical model described above is used to study the response of the different InGaAs-based MSM photodetector designs shown in Figs. 1-3. The dark current, breakdown volt-

age, and the speed of response of one unit cell of these devices are numerically calculated and compared. The simulated finger widths and finger spacings for all the MSM's, except where noted otherwise, are $1 \mu\text{m}$ and $3 \mu\text{m}$, respectively. The total thickness for all these devices is $4 \mu\text{m}$ and the n-InGaAs active region thickness is $2 \mu\text{m}$. The wavelength used in this study is $1.55 \mu\text{m}$, suitable for use in the low-loss window of silica based optical fibers. The relevant parameters for the three different materials used in the simulation are summarized in Table I.

A. InAlAs-InGaAs-InP Device

The device of Fig. 1 consists of a 200-Å thick InAlAs layer, lattice-matched to InGaAs, to increase the relatively low Schottky barrier height ($\sim 0.2 \text{ eV}$) of InGaAs. The metal

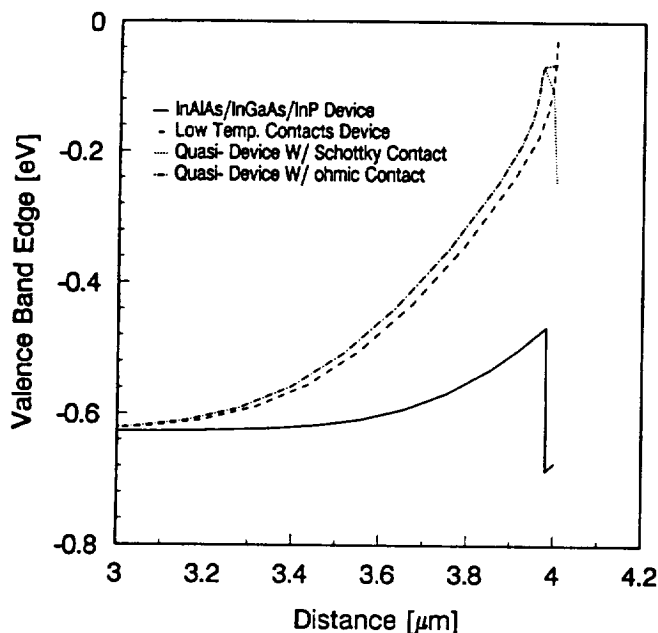


Fig. 5. A one-dimensional slice of the two-dimensional equilibrium solution of the valence band edge diagram for the devices of Figs. 1–3. Only the top 1 μm (see Fig. 1) of the region directly under the top Schottky contacts is shown for clarity. Notice the abrupt hole barrier for the device of Fig. 1 (solid curve) and the soft hole barrier for the device of Fig. 2 with true Schottky contacts (dotted curve.)

Schottky fingers deposited on the InAlAs layer are assumed to form an electron barrier height of 0.65 eV. The doping density in each of the 3 layers of Fig. 1 is taken as 10^{15} cm^{-3} . The top InAlAs layer introduces electron and hole barriers at the InAlAs–InGaAs heterointerface. The hole barrier for the device of Fig. 1 is clearly shown in the equilibrium valence band edge diagram in Fig. 5. For the sake of clarity and ease of comparison, Fig. 5 only shows a one-dimensional slice under the Schottky contacts of the calculated two-dimensional valence band edge diagrams for the MSM devices of Figs. 1–3. From inspection of Fig. 5, we note that for the MSM device of Fig. 1 the InAlAs–InGaAs heterointerface introduces an abrupt hole barrier of 0.21 eV. The electron barrier at the heterointerface is determined as the difference in the affinities of InGaAs and InAlAs (see Table I) and is found to be 0.5 eV.

Assuming no illumination, the electron current flowing toward the collecting Schottky contacts of the device in Fig. 1 is the majority current carrier since the carrier concentration for holes is very small in n-type materials. Consequently, the combination of the high Schottky contact barrier height of the InAlAs top layer and the conduction band edge discontinuity at the InGaAs–InAlAs heterointerface has a great impact on limiting the dark current of the device. The calculated dark current per unit length of the MSM fingers for the InAlAs top layer device structure is plotted against the applied voltage and is shown as the solid line in Fig. 6. The dark currents for the other MSM devices sketched in Figs. 2 and 3 are also plotted in Fig. 6. From inspection of Fig. 6, it is clear that the dark currents per unit length of the MSM fingers for the different devices considered here are comparable and limited to very small values under low applied bias conditions.

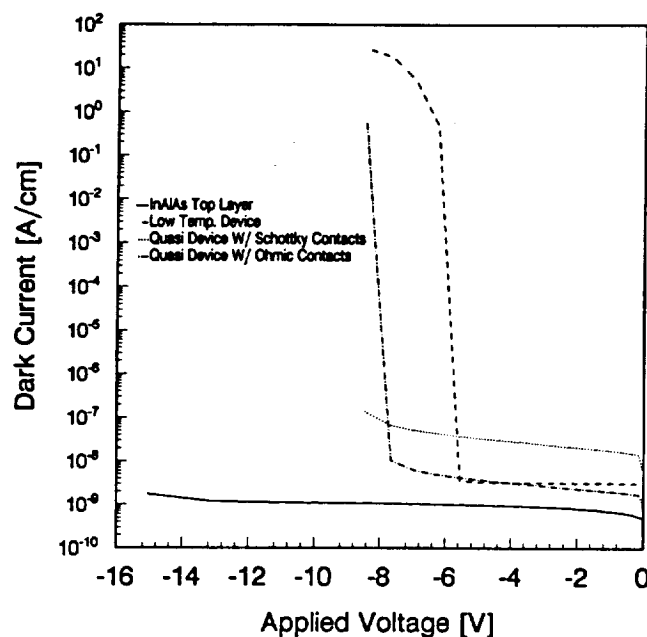


Fig. 6. Dark current per unit length of the MSM fingers vs. applied bias for the devices shown in Figs. 1–3.

As the applied bias is increased, the breakdown voltages for the different devices in Figs. 1–3 are seen to be significantly different. In the device structure in which the top layer is made from InAlAs, no reverse bias breakdown occurs over the applied voltage range considered here. In contrast, very sharp breakdown occurs at nearly the same applied voltages in the other three structures. In these three device structures the contacts are all formed on InGaAs, either using a low temperature growth technique or through use of the p^+ assist layer. The simulator predicts that carrier multiplication through impact ionization occurs primarily in the vicinity of the Schottky contacts where the local electric field reaches its maximum value. Depending upon the applied bias, the local electric field near the Schottky contacts can become sufficiently large to initiate impact ionization. These large fields near the contacts and the subsequent initiation of the impact ionization processes have been reported in GaAs MESFET studies as well [29], where it was found that the impact ionization near the drain and gate edges dominates the transport in these devices. The field distribution near the Schottky contacts can, therefore, be responsible for the slight variation in the breakdown voltages for the devices with an InGaAs layer located under the Schottky contacts.

The impact ionization rate within the underlying layers under the Schottky contacts depends strongly upon the nature of the material composition. In the devices in which the top contacts are made to InGaAs, the breakdown voltages are much lower than in the InAlAs top layer device. This follows directly from the fact that the band gap energy in InGaAs is much lower than in InAlAs and as such, the impact ionization rate is significantly less in InAlAs than in InGaAs at comparable electric field strengths. In Table I, the impact ionization rate parameters for InGaAs and InAlAs used in the simulation are listed for comparison. Further inspection

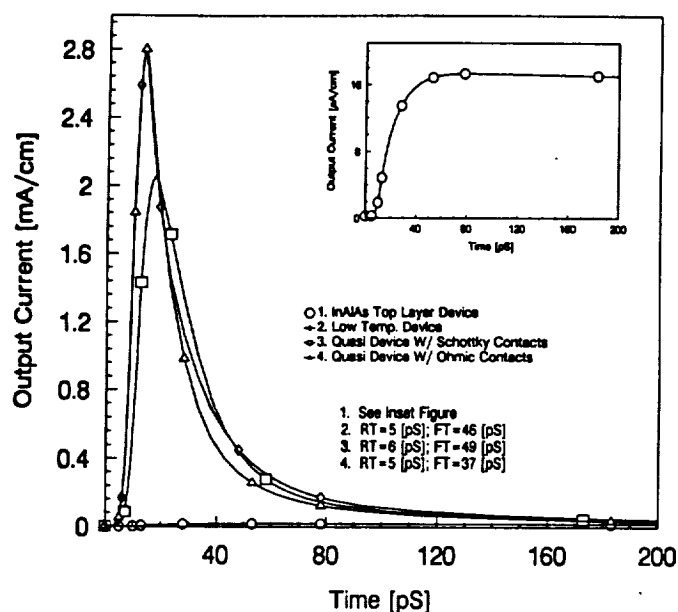


Fig. 7. Calculated current impulse response per unit length of the MSM fingers for the devices shown in Figs. 1–3 at an applied bias of -5 V. Notice that the collected output current for the device of Fig. 1 (circled legend) is very small in comparison to that of the other devices and the shape of the Gaussian input signal is completely lost. The inset of this figure is presented for clarity.

of Fig. 6 shows that the breakdown voltages for the devices that have an InGaAs top layer under the contacts are much lower than that for the device where InAlAs is grown under the Schottky contact. The difference in the impact ionization rate coefficients explains the large difference in the breakdown voltages between the devices with InGaAs and InAlAs layers under the Schottky contacts.

Under illumination conditions, the holes from the optically generated electron-hole pairs (EHP's) become the majority current carriers collected at the Schottky contacts of the device in Fig. 1. The valence band edge discontinuity from the InAlAs–InGaAs heterointerface can act to drastically limit the collected signal current at the Schottky contacts. Only holes that have sufficient kinetic energy to thermionically emit over the hole barrier or tunnel through it can contribute to the signal current collected at the Schottky contacts. The calculated current impulse response for the InAlAs top layer device collected at the biased Schottky contact is shown as the circled legend in Fig. 7, and repeated in the inset of the same figure for clarity. The input power density used is 1 mW/cm^2 and the absorption coefficient of the n-InGaAs active layer at $1.55 \text{ } \mu\text{m}$ is $8 \times 10^3 \text{ cm}^{-1}$ [20]. The input signal is assumed gaussian in time and has a full width at half maximum (FWHM) of 5 ps. The applied bias is held constant at -5 V. It is evident from Fig. 7 that there is almost no collection of the current signal at the biased Schottky contact. Moreover, the shape of the input signal is completely lost. This is partly due to the relatively high valence band edge discontinuity at the InAlAs–InGaAs heterointerface and to the fact that the carrier energies are fixed to the lattice thermal energy within the context of the drift-diffusion model. As a result little appreciable hole current flows through the top contacts. Instead, most of the current flow is through the substrate.

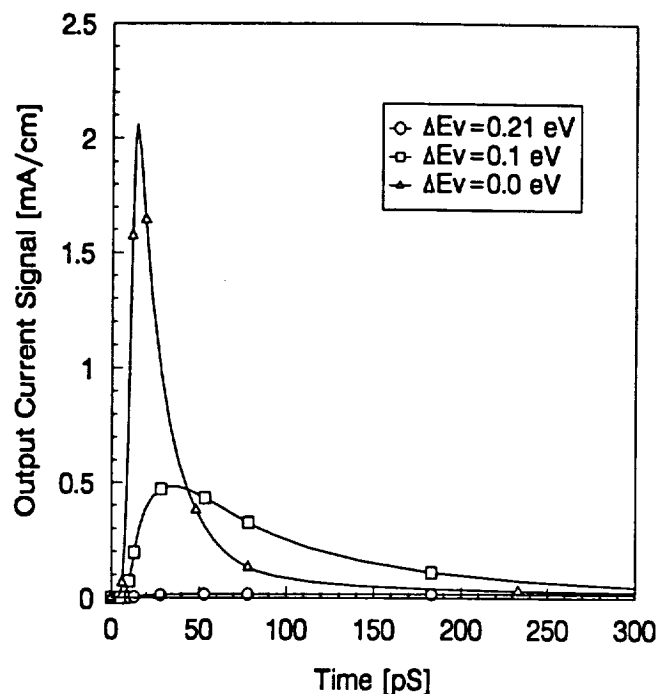


Fig. 8. Calculated current impulse response per unit length of the MSM fingers for the InAlAs top layer device of Fig. 1 at an applied bias of -5 V as the valence band edge discontinuity at the InAlAs–InGaAs heterointerface, ΔE_v , is varied from 0.21 eV to 0.0 eV. Notice the sharp increase in the output current peak as ΔE_v approaches 0 eV.

To demonstrate the effect of the valence band edge discontinuity, we fictitiously changed the affinity of InAlAs in the simulation in such a way as to eliminate the hole barrier at the InAlAs–InGaAs heterointerface. Fig. 8 shows the calculated current impulse response with ΔE_v set to zero at the InGaAs–InAlAs heterointerface. For the sake of comparison, Fig. 8 also shows the impulse response with the correct valence band edge discontinuity of 0.21 eV and a median value of 0.1 eV. As can be seen from Fig. 8, the collected current signal increases dramatically as the valence band edge discontinuity decreases toward zero. Moreover, the output signal starts to approach the gaussian shape of the input as the valence band edge discontinuity approaches zero. On the basis of these observations, it appears that the hole barrier at the InAlAs–InGaAs heterointerface significantly limits the collected current signal at the Schottky contacts of the device.

In addition to thermionic emission at the heterointerface, carrier tunneling can also occur. The effect of tunneling at the heterointerfaces of the device on the collected current is examined using the WKB approximation for calculating the tunneling probability and the Boltzmann distribution function for calculating the tunneling current [15]. In this model, the tunneling current is incorporated as a boundary condition at the heterojunction interfaces [15] and is added to the limiting current obtained from the drift-diffusion/thermionic emission models as described above. Fig. 9 shows the calculated impulse response for this device with and without the tunneling model. Also shown in Fig. 9 is the impulse response of the device as the top InAlAs layer thickness is varied. Notice that as the InAlAs layer thickness is decreased, which in turn

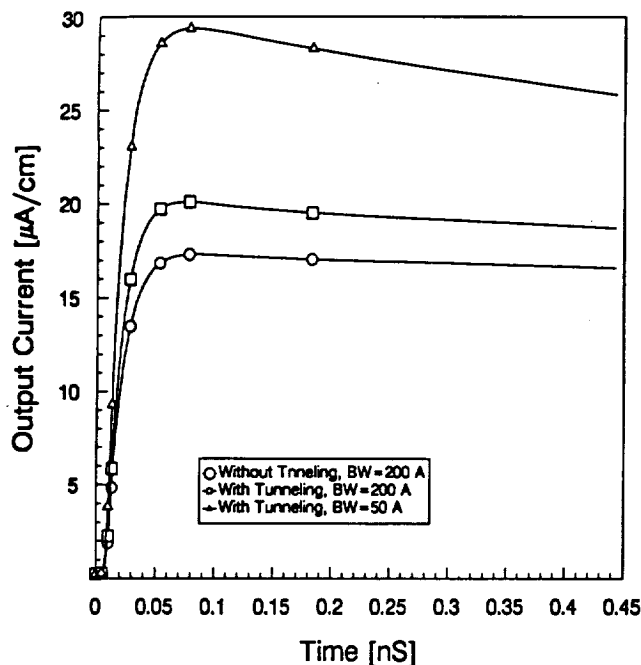


Fig. 9. Current impulse response per unit length of the MSM fingers for the InAlAs top layer device of Fig. 1 at an applied bias of -5 V calculated with (square) and without (circles) the tunneling current for an InAlAs barrier width (BW) of 200 Å. The curve with the triangular legend is the response (including tunneling) of the same device, but with an InAlAs BW of 50 Å.

decreases the width of the barrier, the tunneling contribution to the output current signal increases. In addition, for the different InAlAs layer thicknesses shown in Fig. 9 the effect of the tunneling current on the collected current signal is clearly negligible. Moreover, the gaussian shape of the input signal is not clearly observed in the collected output current signal even when tunneling is included into the model.

In an effort to explain the temporal response of the device in Fig. 1, we compared results obtained from the present model to the experimental results by the authors in [5], [30] and the Monte Carlo studies of Sano [31]. Our model predicted extremely long fall times in comparison to the results reported in these references for devices that are similar in geometry and operating conditions. A possible explanation for this discrepancy lies within the assumptions of the drift diffusion model itself. Contrary to the Monte Carlo model of Sano [31], the drift diffusion model assumes that carriers lie at the edges of the bands in terms of energy and it does not account for any hot carrier transport. Furthermore, carrier energies are fixed to the lattice thermal energy. This is not always true especially under high field variations or small finger spacings of MSM's where carrier energies are much higher than the lattice thermal energy. Consequently, the drift diffusion model underestimates the actual transport across the InAlAs-InGaAs heterojunction because the distribution of carrier energies is assumed to be in equilibrium with the lattice. A more reasonable distribution function would be a heated distribution that takes into account hot carrier transport. Therefore, to accurately describe the transport for this device a more advanced model capable of tracking the energy of the carriers and allowing for hot carrier effects may be necessary. Further investigations of the

device studied here will be pursued in a future work using an advanced energy balance model.

The blocking nature of the hole barrier at the abrupt InGaAs-InAlAs heterointerface can be minimized by using a graded $\text{In}_{0.53}(\text{Ga}_x\text{Al}_{1-x})_{0.47}\text{As}$ layer between the absorbing InGaAs layer and the InAlAs barrier layer. This idea has been investigated by several authors [30]–[32] in which they compared the response of graded and abrupt junction devices and found that the use of a graded $\text{In}_{0.53}(\text{Ga}_x\text{Al}_{1-x})_{0.47}\text{As}$ layer/superlattice between the InGaAs and the InAlAs yields a faster response than just using an abrupt junction. Although the use of a graded layer may enhance the performance of InGaAs based MSM's, it deprives them from their distinct attributes of simplicity and compatibility with existing FET technologies.

B. P^+ Assist Layer, Quasi-Schottky Device

It has been reported [9] that the Schottky barrier height on InGaAs can be effectively increased by growing a thin layer of p^+ -InGaAs on top of an n-InGaAs layer. A schematic diagram of such a structure is shown in Fig. 2. The n^+ InGaAs layer, whose function is to prevent rectification at the InGaAs/InP junction, is chosen to be $0.2 \mu\text{m}$ thick and doped at 10^{18}cm^{-3} . The n^+ InP buffer layer is taken as $0.1 \mu\text{m}$ thick and is also doped at 10^{18}cm^{-3} . The top p^+ InGaAs layer thickness and doping density have to be designed in such a way that this layer is fully depleted and the barrier height enhancement is maximized [9]. The layer thickness used in this paper is 36.5 nm and the doping density is 10^{18}cm^{-3} . The barrier height for the Schottky contacts deposited on the p^+ InGaAs layer is -0.18 eV [9]. Since the conduction and valence bands for this device track each other (no heterojunctions present), Fig. 5 shows that the effective electron barrier height is ~ 0.54 eV. This effective barrier height is measured as the difference in potential energy between the potential maximum and the edge of the depletion region (see Fig. 5.) This relatively high effective barrier height limits the dark current of the device to acceptably small values. A plot of the dark current as a function of voltage is shown with the dotted legend in Fig. 6. Note that the impact ionization breakdown voltage for this device occurs at a much lower voltage than that of the device of Fig. 1. This low breakdown voltage is due to impact ionization near the Schottky contact edges where the local electric field is very large and to the high impact ionization rate coefficients of the low-bandgap InGaAs material.

From the band edge diagram shown in Fig. 5, it is clear that the p^+ InGaAs layer is fully depleted and that the Schottky contacts form a soft barrier for holes. This hole barrier will limit the collection of the photogenerated holes at the Schottky contacts. This can easily be seen by examining the impulse response of this device and comparing it to the response of the same device with the exception that the Schottky contacts are replaced with ohmic contacts. Fig. 7 shows the output current signal response per unit length of the MSM finger as a function of the collection time for this device with Schottky and ohmic contacts. Since there are practically no barriers for holes under the ohmic contacts (See Fig. 5), the output signal

current collected at the ohmic contacts is clearly much higher than that collected at the Schottky contacts. Fig. 7 also shows the rise and fall times of the different devices studied. It is clear that the impulse response for the quasi-Schottky device of Fig. 2 with ohmic contacts outperforms that of the same device with true Schottky contacts. However, it is much harder to grow ohmic contacts on heavily doped p-type materials.

C. Low Temperature-Grown Schottky Contact Device

Recently it has been reported [10] that the barrier height on n-InGaAs can also be improved by growing the Schottky contacts at cryogenic temperatures. Lee *et al.* [10] achieved a barrier height of 0.6 eV by growing Ag metal Schottky contacts at a temperature of 77 K. The device of Fig. 3 assumes the barrier height of the Schottky contacts is 0.6 eV. The InGaAs and the InP layers are 2 μm thick each and doped n-type at 10^{15}cm^{-3} . The Schottky barrier height of this device, which determines the dark current response, is comparable to those for the devices of Figs. 1 and 2. Fig. 6 shows the dark current as a function of the applied bias for this device. Note that the breakdown voltage occurs at ~ 5.5 V. This low breakdown voltage, again, is due to the combination of high impact ionization rate coefficients in InGaAs and the large electric fields near the Schottky contact edges.

The time impulse response of this device is studied next. From inspection of the valence band diagram shown as the dashed curve in Fig. 5, it is clear that the valence band edge in this device has no hole barrier under the Schottky contacts. This fact is very attractive since it leads to a large collection efficiency of the photogenerated holes at the Schottky contacts as in the case of the p^+ assist layer device of Fig. 2 with ohmic contacts. Yet the electron dark current is greatly reduced by the relatively large Schottky barrier height. The collected output signal current per unit length of the MSM fingers of the low temperature grown Schottky contact device is shown as the curve with the diamond legend in Fig. 7. Notice that the output signal current peaks at roughly the same value as that found for the p^+ assist layer device with ohmic contacts. Both of these structures show a significantly higher peak in the output signal current in comparison to the peak response of the other MSM's studied. Additionally, both structures show a much smaller full width at half maximum than either the InAlAs top layer device or the p^+ assist layer device with Schottky contacts. Fig. 7 also shows the rise and fall times of the low temperature grown Schottky contact device. The high responsivity of this device is due mainly to the absence of hole barriers under the collecting Schottky contacts.

IV. CONCLUSION

Excessively high dark currents arise in metal-semiconductor-metal InGaAs based photodetectors due to the low Schottky barrier heights typically formed on this material. Several different approaches have been suggested to reduce the dark currents by effectively enhancing the potential barrier heights for electrons at the collecting contacts in InGaAs based MSM devices. In this paper, we have theoretically examined the response of three different

InGaAs-based interdigitated MSM photodetectors, suitable for use in the low loss window of silica optical fibers, which incorporate different schemes for increasing the potential barrier height and thereby reducing the dark current. The calculations are made using a two-dimensional drift diffusion model that incorporates both thermionic emission and tunneling across heterobarriers.

The first device structure, that shown in Fig. 1, where an InAlAs layer with a Schottky barrier height of ~ 0.65 eV is grown on top of the active InGaAs layer. This device enhances the effective Schottky barrier height but it introduces considerable electron and hole barriers at the InGaAs-InAlAs heterointerface. The effective barrier for electrons acts to limit the dark current of this device. The hole barrier, however, drastically limits the collection of the photogenerated holes from the InGaAs active region of the device when operated at low to moderate biases. We have shown that, in the absence of the hole barrier at the InGaAs-InAlAs heterointerface the collected current signal at the top Schottky contacts of this device is much higher than the case when this barrier is present and the output signal retains the shape of the gaussian input signal. The tunneling current contribution across the 200 Å wide square barrier formed under the Schottky contacts is found to be negligibly small to affect the collected output current signal at the applied bias of -5 V used here. As the InAlAs barrier layer thickness is made smaller the tunneling current contribution starts to increase. It is expected that at higher applied biases the voltage drop in the barrier region increases and the barrier starts to look triangular in shape where a more appreciable tunneling current contribution is possible. However, for InAlAs barrier layer thicknesses of 200 Å or more, as has been used for similar devices in the literature [5], [6], [30]–[32], and at an applied bias of -5 V, the calculated tunneling current is found to have little effect on the collection of the photogenerated holes. We have indicated that the drift diffusion model may be insufficient to describe the actual transport across the InAlAs-InGaAs heterojunction and that a more advanced, energy balance, model is necessary. The distinct features of this device, nevertheless, are the facts that the dark current is very small, as is clear from Fig. 5, and the breakdown voltage occurs at a much higher bias than that of the devices of Figs. 2 and 3.

An alternative means of enhancing the effective Schottky barrier height on InGaAs, using a p^+ assist, quasi-Schottky device shown in Fig. 2 is considered. For this device the top p^+ InGaAs assist layer forms a p^+ -n homojunction with the active n-InGaAs layer. The depletion region of these two layers forms what is called a quasi-Schottky contact with a barrier height of ~ 0.54 eV. However, the Schottky contacts deposited on the p^+ InGaAs layer introduce a soft hole barrier (See Fig. 5) that acts to limit the collection efficiency of this device. This hole barrier can be eliminated by depositing ohmic contacts at the p^+ InGaAs top layer instead of the existing Schottky contacts. The transient response of this device with ohmic contacts is very attractive since high responsivity and speed of response are achieved as is evident from Fig. 7. However, the disadvantage of using ohmic

contacts in this device lies in the fact that it is generally much harder to make ohmic contacts on p^+ materials. The quasi-Schottky device with ohmic or Schottky top contacts suffers from low breakdown voltages in comparison with the InAlAs top layer device of Fig. 1. Moreover, many more processing steps are required for this device during growth.

Finally, the response of the device in Fig. 3 where a barrier height of 0.6 eV was obtained by depositing Schottky contacts at a temperature of 77 K is examined. The calculated dark current of this device is found to be very small and comparable to that of the quasi-Schottky device with ohmic contacts. The breakdown voltage for this device is rather poor and it is due to the large local electric fields near the Schottky contact edges and the low bandgap of InGaAs. A high responsivity is obtained for this device as is shown in the calculation of the output current impulse response in Fig. 7. This high responsivity is mainly due to the absence of any hole barriers under the Schottky contacts since holes are the majority current carriers at the top Schottky contacts under illumination conditions. Note that the peak in the output signal response of this device is practically the same as that of the p^+ assist, quasi-Schottky device with ohmic contacts since for both of these devices there are no hole barriers present under the collecting contacts. The simplistic structure of this device is advantageous over those of Figs. 1 and 2 since fewer processing steps are required. However, the stability of the low-temperature grown Schottky contacts as a function of operating temperatures is not fully established. More experimental work is needed to study the effect of temperature variations on the barrier height and the response of this device.

Based on the calculations presented here, optimal performance as measured in terms of limited dark current and speed of response of InGaAs based MSM structures is achievable using either low-temperature grown Schottky contacts or using a p^+ quasi-Schottky device with ohmic contacts. Though these device structures are predicted to outperform other InGaAs MSM designs in terms of speed of response they suffer from low breakdown voltages due to the small bandgap of InGaAs. Furthermore, the full exploitation of these two device structures may be limited by materials issues.

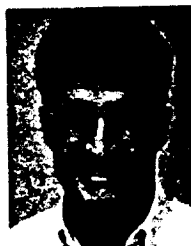
ACKNOWLEDGMENT

The authors would like to thank Dr. A. Smith for many helpful comments during the course of this work. The authors are also indebted to Dr. L. Tarof at BNR for help in providing material parameters used in this work.

REFERENCES

- [1] S. Y. Chou and M. Y. Liu, "Nanoscale tera-hertz metal-semiconductor-metal photodetectors," *IEEE J. Quantum Electron.*, vol. 28, no. 10, pp. 2358-2368, Oct. 1992.
- [2] L. Figueroa and C. W. Slayman, "A novel heterostructure interdigitated photodetector (HIP) with picosecond optical response," *IEEE Electron Device Lett.*, vol. EDL-2, no. 8, pp. 208-210, Aug. 1981.
- [3] M. Ito and O. Wada, "Low dark current GaAs metal-semiconductor-metal (MSM) photodiodes using WSi_2 contacts," *IEEE J. Quantum Electron.*, vol. QE-22, pp. 1073-1077, 1986.
- [4] K. Nakajima, T. Iida, K. Sugimoto, H. Kan, and Y. Mizushima, "Properties and design theory of ultrafast GaAs metal-semiconductor-metal photodetector with symmetrical Schottky contacts," *IEEE Trans. Electron Devices*, vol. 37, no. 1, pp. 31-35, Jan. 1990.
- [5] J. B. D. Soole and H. Schumacher, "InGaAs metal-semiconductor-metal photodetectors for long wavelength optical communications," *IEEE J. Quantum Electron.*, vol. 27, no. 3, pp. 737-752, Mar. 1991.
- [6] J. H. Burroughes and M. Hargis, "1.3 μm InGaAs MSM photodetector with abrupt InGaAs/AlInAs interface," *IEEE Photon. Technol. Lett.*, vol. 3, no. 6, pp. 532-534, June 1991.
- [7] J. B. D. Soole and H. Schumacher, "Transit-time limited frequency response of InGaAs MSM photodetectors," *IEEE Trans. Electron Devices*, vol. 37, no. 11, pp. 2285-2291, Nov. 1990.
- [8] C. Shi, D. Grützmacher, M. Stollenwerk, Q. Wang, and K. Heime, "High-performance undoped InP/n-In_{0.53}Ga_{0.47}As MSM photodetectors grown by LP-MOVPE," *IEEE Trans. Electron Devices*, vol. 39, no. 5, pp. 1028-1031, May 1992.
- [9] S. V. Averin, A. Kohl, R. Müller, J. Wisser, and K. Heime, "Quasi-Schottky barrier MSM-diode on n-In_{0.53}Ga_{0.47}As using a depleted p^+ -In_{0.53}Ga_{0.47}As layer grown by LP-MOVPE," *Solid-State Electron.*, vol. 36, no. 1, pp. 61-67, 1993.
- [10] H. J. Lee, W. A. Anderson, H. Hardtdegen, and H. Lüth, "Barrier height enhancement of Schottky diodes on n-In_{0.53}Ga_{0.47}As by cryogenic processing," *Appl. Phys. Lett.*, vol. 63, no. 14, pp. 1939-1941, Oct. 1993.
- [11] A. Salem, A. Smith, and K. Brennan, "Theoretical study of the effect of an AlGaAs double heterostructure on metal-semiconductor-metal photodetector performance," *IEEE Trans. Electron Devices*, vol. 41, no. 7, pp. 1112-1119, July 1994.
- [12] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. New York: Springer-Verlag, 1984.
- [13] F. Capasso, "Physics of avalanche photodiodes," in *Semiconductors and Semimetals*, vol. 22, Part D, R. K. Willardson and A. C. Beer, Eds. New York: Academic Press, 1985, pp. 1-172.
- [14] K. Horio, and H. Yanai, "Numerical modeling of heterojunction including the thermionic mechanism at the heterojunction interface," *IEEE Trans. Electron Devices*, vol. 37, no. 4, pp. 1093-1098, Apr. 1990.
- [15] K. Yang, J. R. East, and G. Haddad, "Numerical modeling of abrupt heterojunctions using a thermionic-field emission boundary condition," *Solid-State Electron.*, vol. 36, no. 3, pp. 321-330, 1993.
- [16] J. D. McGervey, *Introduction to Modern Physics*. New York: Academic Press, 1971.
- [17] D. L. Scharfetter, and H. K. Gummel, "Large-Signal analysis of a silicon read diode oscillator," *IEEE Trans. Electron Devices*, vol. ED-16, no. 1, pp. 64-77, Jan. 1969.
- [18] G. H. Golub, C. F. Van Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins Univ. Press, 1983, pp. 362-379.
- [19] T. C. Oppe, W. D. Joubert, and D. R. Kinkaid, "An overview of NSPCG: A nonsymmetric preconditioned conjugate gradient package," *Comput. Phys. Commun.*, vol. 53, pp. 283-293, 1989.
- [20] K. Yasuda, T. Shirai, Y. Kishi, S. Yamazaki, and T. Kaneda, "Heterojunction effects on spectral and frequency responses in InP/InGaAsP/InGaAs APD," in *Proc. 14th Int. Conf. on Solid-State Devices*, Tokyo, Japan, 1982. (Also in *Jap. J. App. Phys.*, vol. 22, Suppl. 22-1, pp. 291-294, 1983.)
- [21] K. Brennan, "Theoretical study of multiquantum well avalanche photodiodes made from the GaInAs/AlInAs material system," *IEEE Trans. Electron Devices*, vol. ED-33, no. 10, pp. 1502-1510, Oct. 1986.
- [22] E. H. C. Parker, Ed., *The Technology and Physics of Molecular Beam Epitaxy*. New York: Plenum Press, 1985.
- [23] A. Elshabini-Riad and J. He, "Investigation of injection mechanisms for InGaAs/InP double heterostructure bipolar transistors," *Solid-State Electron.*, vol. 32, no. 10, pp. 853-860, 1989.
- [24] S. Hausser, G. Fuchs, A. Hangleiter, and K. Streubel, "Auger recombination in bulk and quantum well InGaAs," *Appl. Phys. Lett.*, vol. 56, no. 10, pp. 913-915, Mar. 1990.
- [25] S. Adachi, "Material parameters of $In_{1-x}Ga_xAs_yP_{1-y}$ and related binaries," *J. Appl. Phys.*, vol. 53, no. 12, Dec. 1982.
- [26] K.-H. Hellwege and O. Madelung, Eds., *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology*. New York: Springer-Verlag, 1982.
- [27] G. Augine, "Modeling, fabrication, and characterization of InP thin films and devices for optoelectronic applications," Ph.D. dissertation, Georgia Inst. of Technol., Atlanta, GA, 1994.

- [28] M. Takeshima, "Simple method of calculating phonon-assisted Auger recombination rate in direct-gap semiconductor," *J. Appl. Phys.*, vol. 22, no. 3, pp. 491-498, 1983.
- [29] C.-L. Li, T. M. Barton, and R. E. Miles, "Avalanche breakdown and surface deep-level trap effects in GaAs MESFET's," *IEEE Trans. Electron Devices*, vol. 40, no. 4, pp. 811-816, Apr. 1993.
- [30] E. Sano, M. Yoneyama, T. Enoki, and T. Tamamura, "Performance dependence of InGaAs MSM photodetectors on barrier-enhancement layer structures," *Electron. Lett.*, vol. 28, no. 13, pp. 1220-1221, June 1992.
- [31] E. Sano, "Theoretical analysis of the influences of barrier-enhancement layers on transient responses of MSM photodetectors," *IEEE Trans. Electron Devices*, vol. 39, no. 6, pp. 1355-1362, June 1992.
- [32] O. Wada, H. Nobuhara, H. Hamaguchi, T. Mikawa, A. Tackeuchi, and T. Fujii, "Very high speed GaInAs metal-semiconductor-metal photodiode incorporating an AlInAs/GaInAs graded superlattice," *Appl. Phys. Lett.*, vol. 54, no. 1, pp. 16-17, Jan. 1989.



Ali F. Salem (S'90) was born in Marabon, Lebanon, on September 2, 1967. He received the B.S. degree from Boston University, Boston, MA, in 1989, and the M.S. degree from the Georgia Institute of Technology, Atlanta, GA, in 1990.

He is currently finishing his Ph.D. degree in electrical engineering at the Georgia Institute of Technology. His dissertation work involves drift diffusion and hydrodynamic modeling of carrier transport in semiconductor devices and its applications to metal-semiconductor-metal photodetectors.

Kevin F. Brennan (S'84-M'84-SM'90), photograph and biography not available at the time of publication.

Influence of Hot Carrier Transport on the Transient Response of an InGaAs/InAlAs Metal-Semiconductor Schottky Diode Structure

Ali F. Salem and Kevin F. Brennan

Abstract—The calculated transient characteristic of a heterostructure, rectifying contact is theoretically examined. It is found that hot carrier transport drastically affects the output terminal characteristics of the heterostructure Schottky contact and, hence, the workings of a blocking contact. This is of importance to the understanding of InGaAs MSM devices in particular, as well as any structure which contains a blocking contact in general.

In this paper, we use a HD model [1] that is capable of tracking the carrier energies as well as the carrier concentrations and carrier velocities to study the temporal response of the current in a blocking, heterostructure contact. Specifically, we examine an important example of a blocking contact, an InGaAs/InAlAs Schottky diode structure that is commonly present in InGaAs-based metal-semiconductor-metal (MSM) photodetectors to reduce the dark current.

Previous theoretical analysis of InGaAs-based MSM's [2] using a drift-diffusion (DD) model showed that the top InAlAs layer introduces electron and hole barriers near the metal Schottky contacts of the MSM that limit not only the dark current of the structure but its transient response as well. Transient response calculations using the DD model predicted extremely long rise and fall times [2] but these predictions seem to be inconsistent with measured data on comparable device structures [3]. In this work, we will show using HD calculations of the transient response of this device, that carrier heating effects can result in much shorter rise and fall times than those predicted using the DD model which are more consistent with experiment.

The complete hydrodynamic (HD) model used in this work is an extension of the DD model [2], [4] which includes the additional solution of the electron and hole carrier energies based on the approach of Bløteckjaer [1] and Azoff [5]. Analytical models for the mobilities and constant values for the relaxation times are utilized since more accurate Monte Carlo based parameters are not presently available to us.

The HD system of equations is solved in series with a thermionic emission boundary condition for Schottky contacts and heterojunctions. Only carriers that have sufficient kinetic energy to surmount the barriers at a heterojunction interface are emitted from one side of the junction to the other. The thermionic emission current condition is based on Crowell's and Sze's thermionic emission-diffusion theory [6]. The extension of this theory to hot electron transport has been derived by several authors [7]–[9]. The thermionic emission boundary conditions for electrons and holes at heterojunction interfaces implemented in this work are similar to those reported in [9].

Manuscript received July 24, 1995; revised October 30, 1995. The review of this brief was arranged by Editor Alan H. Marshak. This work was supported in part by ARPA through Contract to NASA, NAGW-2753, and by the Digital Equipment Corp., through Contract E21-H36.

The authors are with the School of Electrical Engineering and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, GA 30332 USA.

Publisher Item Identifier S 0018-9383(96)02690-1.

The HD model is used to investigate the temporal response of a structure containing a partially blocking heterostructure Schottky contact. For purposes of illustration and because of its relevance to the study of InGaAs based MSM devices, specifically an InGaAs/InAlAs diode structure with a top Schottky contact is investigated. The doping levels in the semiconductor layers are fixed to 10^{15} cm^{-3} *n*-type. The top InAlAs layer is 80 nm thick and the thickness of the active InGaAs layer is $1.92 \mu\text{m}$. The width of the device is assumed to be $2.0 \mu\text{m}$. The material parameters used in the simulation are the same as those used in [2]. The input optical generation signal is gaussian in time and decays exponentially away from the top of the device. The full width at half maximum (FWHM) of the input signal is 5 ps and the input power is 1 mW/cm^2 . For simplicity, the top Schottky contact is assumed to be transparent. The wavelength of the input signal is taken as $1.55 \mu\text{m}$.

To examine the effect of carrier heating in the device, the carrier energies for both electrons and holes are calculated at an applied bias of -15 V . These results are shown in Fig. 1. From inspection of Fig. 1 it is clear that the electrons and holes are significantly heated beyond thermal equilibrium conditions. Provided the carrier heating is sufficient to give the carriers enough energy to offset the effect of the heterojunction barrier, then the transport near the heterojunction interface can change significantly and the long rise and fall time of the transient response predicted using the DD model will, in turn, be significantly diminished. As can be seen from Fig. 1, the hole energy drops dramatically near the $2 \mu\text{m}$ length end of the device due to the presence of the InGaAs/InAlAs heterobarrier. Recall that the holes move from left to right in the diagram (the negative voltage is applied to the right hand contact) while the electrons move in the opposite direction. After crossing the heterobarrier, the holes are first cooled by the loss of kinetic energy to the potential barrier but are then slightly heated by the field within the InAlAs layer until they are collected at the Schottky contact. The holes are collected at the Schottky contact at an energy above thermal equilibrium. The electrons, on the other hand, are injected from the Schottky contact into the InAlAs layer. They then cross the InAlAs/InGaAs heterobarrier gaining kinetic energy from the potential discontinuity. The $x = 0$ boundary is ohmic and therefore the electric field vanishes at that point. As such both the electron and hole energies essentially thermalize. However, the thermalization of electron and hole energies occur at different rates and at different locations in the structure. This is possibly due to the fact that the holes are injected into an accelerating field while the electrons encounter a deceleration in the vicinity of the ohmic contact.

Next, the calculated impulse response of the InGaAs/AlInAs Schottky diode device is examined. Fig. 2 shows the calculated impulse response at -15 V using the DD model (solid curve) used previously [2] and the HD model (dashed curve) described above. Close examination of Fig. 2 reveals that the rise and fall times of the output current signal of the HD model are dramatically shorter than those of the DD model and the output signal much more closely resembles the input gaussian signal. Fig. 2 also shows that the responsivity as measured in terms of the maximum of the output current signal is much higher for the HD results than for the DD results. The calculations were repeated using an applied bias of -5 V as shown in the inset of Fig. 2. Again, there is a substantial difference between the HD and DD predicted responses with the HD calculation in much closer agreement with the experiment [3].

To check the importance of other effects, tunneling was added to the DD model following the method of [10]. It was found that

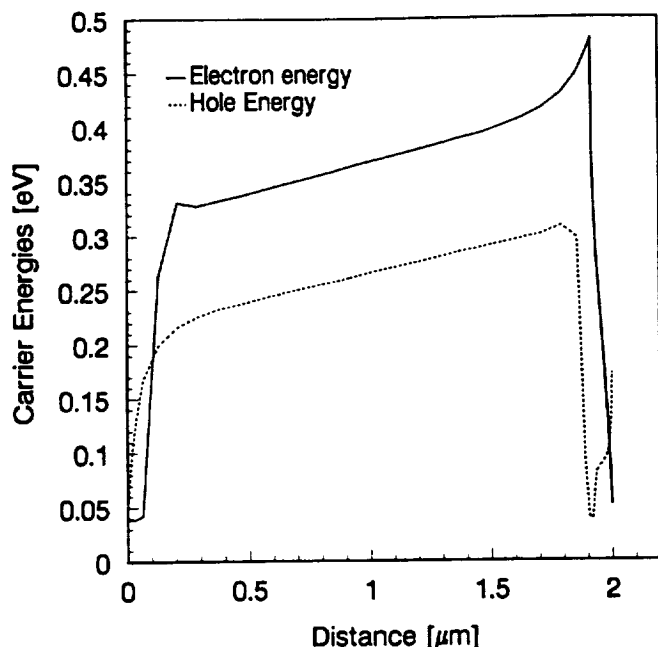


Fig. 1. Electron and hole energy profiles for the InGaAs/InAlAs Schottky diode structure at an applied bias of -15 V. The abrupt change in the energy at the heterojunction interface near the right side contact is due to the limiting thermionic emission boundary condition.

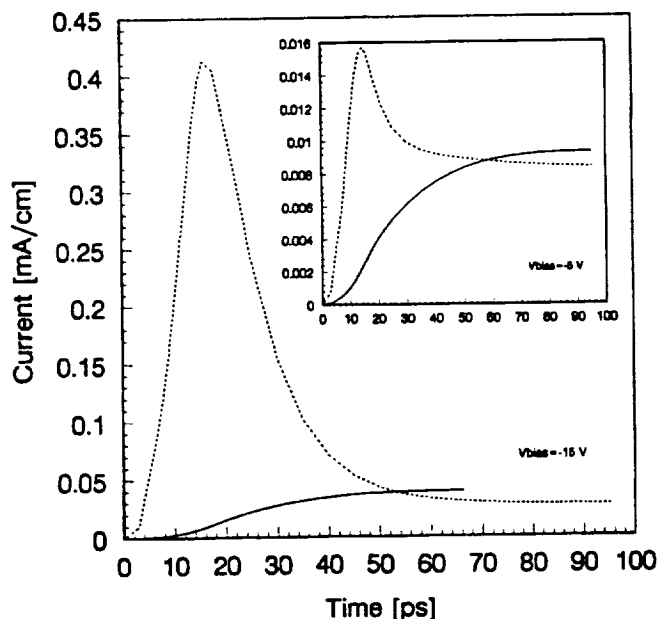


Fig. 2. Calculated current transient response of the device obtained from the DD model (solid curve) and the HD model (dashed curve) at an applied bias of -15 V. Note that the HD model predicts a gaussian shaped output in relatively close agreement with the input signal while the DD response shows an extremely long rise time. The inset shows the calculated transient response of the device using both the DD (solid curve) and HD (dashed curve) models at an applied bias of -5 V. Again, note that the HD model predicts a gaussian shaped output in relatively close agreement with the input signal while the DD model shows a much longer rise time.

even at -15 V bias the current response for the device considered here is not greatly affected by the inclusion of tunneling, resulting in a qualitatively similar temporal response as shown in Fig. 2. Additionally, impact ionization may possibly be important at the bias

considered here. To check its effect, a field-dependent formulation of the impact ionization rate was included within both the HD and DD models. Calculations indicate that the impact ionization rate also does not greatly affect the calculated current response. The DD model still fails to recover the input gaussian signal even when both tunneling and impact ionization are included. This result indicates that the failure of the DD model to reproduce the correct temporal response is most probably due to its failure to track the energies of the carriers.

It is important to note that carrier heating is determined to a large extent by the relaxation times used in the simulation. The relaxation time used in the simulation is a lumped parameter representing the contribution of all of the possible scattering relaxation mechanisms and is simply an estimate of these effects. The relaxation times are not well known for the materials used in this device, and for that reason we examined various values of the relaxation times, from 0.1–0.5 ps, on the calculated transient response. It is found that as the relaxation times decrease, the effect of carrier heating on the temporal response becomes less significant and the HD model results approach that of the DD model as expected. At longer relaxation times, 0.4–0.5 ps, a fast response in the calculated output current is obtained.

In conclusion, the DD model poorly accounts for the temporal response of the current flow across a heterostructure, blocking Schottky contact due to the failure of the DD model to properly track the carrier energies and hence their respective currents. Using a more advanced HD model, which properly tracks the carrier energies, it is seen that much shorter rise and fall times are predicted, in closer agreement with experimental measurements made on an MSM device containing a similar heterostructure Schottky contact.

REFERENCES

- [1] K. Bløtekjaer, "Transport equations for electrons in two-valley semiconductors," *IEEE Trans. Electron Devices*, vol. ED-17, no. 1, pp. 38–47, Jan. 1970.
- [2] A. F. Salem and K. F. Brennan, "Theoretical study of the response of InGaAs metal-semiconductor-metal photodetectors," *IEEE J. Quantum Electron.*, vol. 31, no. 5, pp. 944–953, May 1995.
- [3] J. B. D. Soole and H. Schumacher, "InGaAs metal-semiconductor-metal photodetectors for long wavelength optical communications," *IEEE J. Quantum Electron.*, vol. 27, no. 3, pp. 737–752, Mar. 1991.
- [4] A. F. Salem, A. W. Smith, and K. F. Brennan, "Theoretical study of the effect of an AlGaAs double heterostructure on metal-semiconductor-metal photodetector performance," *IEEE Trans. Electron Devices*, vol. 41, no. 7, pp. 1112–1119, July 1994.
- [5] E. M. Azoff, "Closed-form method for solving the steady-state generalized energy-momentum conservation equations," *COMPEL*, vol. 6, no. 1, pp. 25–30, 1987.
- [6] C. R. Crowell and S. M. Sze, "Current transport in metal-semiconductor barriers," *Solid-State Elect.*, vol. 9, pp. 1035–1048, 1966.
- [7] H. Hjelmgren, "Numerical modeling of hot electrons in n-GaAs Schottky-barrier diodes," *IEEE Trans. Electron Devices*, vol. 37, no. 5, pp. 1228–1234, May 1990.
- [8] D. Schroeder, "The inflow moments method for the description of electron transport at material interfaces," *J. Appl. Phys.*, vol. 72, no. 3, pp. 964–970, Aug. 1992.
- [9] H. Hjelmgren and T.-W. Tang, "Thermionic emission in a hydrodynamic model for heterojunction structures," *Solid-State Elect.*, vol. 37, no. 9, pp. 1649–1657, 1994.
- [10] K. Yang, J. R. East, and G. Haddad, "Numerical modeling of abrupt heterojunctions using a thermionic-field emission boundary condition," *Solid-State Electron*, vol. 36, pp. 321–330, 1993.

Numerical Examination of Silicon Avalanche Photodiodes Operated in Charge Storage Mode

Joseph W. Parks, Jr. and Kevin F. Brennan

Abstract—The behavior of silicon-based avalanche photodiodes (APD's) operated in the charge storage mode is examined. In the charge storage mode, the diodes are periodically biased to a sub-breakdown voltage and then open-circuited. During this integration period, photo-excited and thermally generated carriers are accumulated within the structure. The dynamics of this accumulation and its effects upon the avalanching of the diode warrants a detailed, fully numerical analysis. The salient features of this investigation include device sensitivity to the input photo-current including the self-quenching effect of the diode and its limitations in sensing low light levels, the dependence of the response on the bulk lifetime and hence on the generation current within the device, the initial gain, transient response, dependence of the device uniformity upon performance, and the quantity of storable charge within the device. To achieve these tasks our device simulator, STEBS-2D, was utilized. A modified current-controlled boundary condition is employed which allows for the simulation of the isolated diode after the initial reset bias has been applied. With this boundary condition, it is possible to establish a steady-state voltage on the ohmic contact and then effectively remove the device from the external circuit while still including effects from surface recombination, trapped surface charge, and leakage current from the read-out electronics.

Index Terms—Avalanche photodiodes, numerical simulation.

I. INTRODUCTION

AVALANCHE photodiode (APD's) arrays have received attention as a high-resolution, high-sensitivity alternative to charge-coupled devices (CCD's) for low light level and night vision sensing [1]–[3]. Although it is unclear as to whether these devices will be able to surpass the technologically mature CCD, advances in process uniformity may eventually allow large arrays of these devices to be incorporated in an active pixel sensor (APS) for very high resolution applications [4]. Toward this purpose, the APD is best utilized by configuring it in the charge storage mode where the device can collect charge over the entire integration period rather than just during the short sample/read-out time. In this mode the diode is initially biased to a potential somewhat below the breakdown voltage yet high enough that impact ionization is prevalent, the actual bias being design dependent. Following this reset period the external bias is removed, the diode is physically isolated from the circuit and, over time, recovers

back to equilibrium. The length of time required for the diode to recover is dependent upon a number of parameters such as generation lifetime, quantity of signal and background radiation, and leakage to the read-out circuit. During the first stage of the recovery, the charge added to the diode can be multiplied within the high-field depletion region of the device and is stored within the internal capacitance of the junction. For very low light levels the gain is somewhat uniform throughout the assumed integration period of 10 ms. As a greater number of carriers are generated within the device, however, the stored charge begins to lower the internal bias of the device by filling the depleted space charge region. Hence, the amount of avalanche gain drops in a complex, nonlinear fashion throughout the integration period. This is the previously reported self-quenching mechanism which provides excellent anti-blooming [5]. Furthermore, due to its high sensitivity at low light levels and low gain at high levels, this device provides a response well matched to the sensitivity of the human eye.

Most previous analyses of the charge storage configuration of diodes have relied on analytical approximations to the underlying physical problem [2], [6]. In this study, we numerically examine the operation of the APD using a drift-diffusion simulator which self-consistently solves the Poisson and the current continuity equations. Usage of a completely numerical approach obviates the need for many of the simplifying approximations typically employed. Additionally, nonlinear features such as the field dependent ionization rates and mobilities can be included directly into the simulation.

The primary objective of this work is to provide a general picture of the operation of charge-storage APD's for sensing low light level signals. This is obtained with a light-in to signal-out perspective using a realistic night sky illumination spectrum as an input to the model. The material system of interest here is silicon due to the ease of integration into either APS or CCD technology. Although silicon is particularly inefficient over the wavelengths of interest, it is still the best current option due to the maturity of its technology. Emerging commercial applications such as automotive night time driving aids will most likely rely on silicon-based imagers due to the low cost of the technology and its moderately low dark currents at room temperature.

For comparison, the simulations are performed using two different generic APD designs. First, a reach-through avalanche photodiode (RAPD) design which combines the features of high quantum efficiency and short response time by dividing the voltage drop between a high field multiplication

Manuscript received January 2, 1997; revised August 22, 1997. The review of this paper was arranged by Editor P. K. Bhattacharya. This work was supported in part by DARPA through a contract by NASA, Grant NAGW-2753.

The authors are with the School of Electrical and Computer Engineering and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

Publisher Item Identifier S 0018-9383(98)00941-1.

region and a low field drift region is simulated. Next, a standard p-n junction diode in which the majority of the depletion region supports a high electric field is examined. The response and charge storage capabilities of these two designs differ in the manner in which the depletion region is filled during recovery. Using these designs as a test bed, several numerical experiments are performed. Specifically, we examine the diode sensitivity to various levels of input photocurrent, determine the quantity of storable charge within the structure, observe the dependence of the bulk generation lifetime on the device performance, and provide an analysis of the transient response of the device.

II. MODEL DESCRIPTION

A. Electrical Model

To better understand the performance of APD's operating in the charge storage mode, the sample APD device structures are simulated using the drift-diffusion subset of the Semiconductor Total Energy Balance in Two-Dimension (STEBS-2D) model [7], [8]. In this model, Poisson's equation and the electron and hole continuity equations are solved self-consistently to give the electric potential and the carrier concentrations within the device. It is believed that the drift-diffusion model is sufficient for examining these structures due to the fact that the characteristic dimensions are on the order of microns where the physics of nonstationary transport can be generally neglected. The resulting fundamental partial differential equations used within this study to describe the homogeneous system are

$$\nabla(\epsilon_r \nabla \Psi) = \frac{q}{\epsilon_0} (p - n + N_d^+ - N_a^-) \quad (1)$$

$$-\frac{1}{q} \nabla \cdot J_n - G_n + R_n + \frac{\partial n}{\partial t} = 0 \quad (2)$$

$$\frac{1}{q} \nabla \cdot J_p - G_p + R_p + \frac{\partial p}{\partial t} = 0 \quad (3)$$

$$J_n = q\mu_n n E + qD_n \nabla n \quad (4)$$

$$J_p = q\mu_p p E - qD_p \nabla p \quad (5)$$

where the system variables have their usual representation. The recombinative terms, $R_{n/p}$, include the sum of SRH, Auger and radiative rates [9]. Note that these are predominately generative under reverse bias conditions. The usual field dependent electron and hole mobility models are also used for silicon [9] and the diffusivities follow the Einstein relation.

The generation rates within the device are the important mechanisms which drive the diode toward its equilibrium state. Within this model, these entail the sum of photo-generation and impact ionization rates. The photo-generation rate follows an exponential decay into the crystal given by

$$G_{ph} = \sum_i I(\lambda_i) e^{-(\alpha(\lambda_i)d)} \quad (6)$$

where $I(\lambda)$ is the density of incident photons at wavelength λ and $\alpha(\lambda)$ is the wavelength dependent absorption rate. The impact ionization rate follows the standard field dependent

model

$$G_{ii(n/p)} = \frac{1}{q} |J_{(n/p)}| \alpha_{0(n/p)} e^{-\left(\frac{E_{crit(n/p)}}{E_{||}}\right)^\beta} \quad (7)$$

with the electric field magnitude in the direction of the carrier flux given as $E_{||}$.

The system of equations given by (1)–(5) is nonuniformly discretized using a rectangular control-volume finite-difference method [10]. The resulting system of equations is linearized using a multi-dimensional Newton's method and is solved using the NSPCG mathematics library for nonsymmetric systems [11].

B. Boundary Condition

Typically, when modeling ohmic contacts within a semiconductor device, the boundary condition along the periphery of the simulation domain is mathematically treated as a Dirichlet boundary where the electrostatic potential is set to a constant value for a given time step. This configuration implicitly assumes that the semiconductor device under investigation can be numerically treated as a voltage controlled current source. There are situations, however, when a more appropriate description of the boundary involves specifying the current at the contact rather than the potential. One such application is the examination of the reverse recovery within a diode.

Several simulation techniques which incorporate current-controlled boundaries have been reported [12], [13]. These models rely upon the specification of an external circuit to regulate the current flow out of the contact. In this manner, the potential remains the controlled variable while the carrier concentrations are established using the combined assumptions of charge neutrality and carrier equilibrium at the contact. These techniques are applicable for the transient analysis of a device so long as the displacement current is accounted for within the simulation, the device remains under bias, and the load resistance from the external circuit remains sufficiently low.

Modeling the reverse recovery of an APD, however, involves removing the reset bias and open-circuiting the device. This produces both physical and numerical ramifications which cannot be adequately handled with an external circuit. When the diode is electrically isolated after the initial reset bias has been established, the voltage at the cathode floats. The metal contact is physically disconnected from the infinite source of carriers previously supplied by the external voltage source. Thus, the infinite surface recombination implicit in the standard ohmic boundary is no longer applicable and the assumption of carrier equilibrium at the contact will no longer hold. Additionally, modeling the open-circuited device by using an arbitrarily large load resistance between the contacts produces numerical instabilities arising from the round-off error associated with the calculation of the terminal currents. Furthermore, the cathode voltage is determined solely by the distribution of charge within the diode itself, not by the connection to an external circuit.

Of primary concern is establishing a consistent boundary condition which physically models the recovery process within

the diode. The dissemination of the external current into the respective electron and hole currents at an ohmic contact is a nontrivial problem which is not well defined on the macroscopic level. For simplicity, one method of partitioning the currents is to assume that the total current is distributed into the electron and hole components relative to their respective densities at the contact. This is the approach previously reported in the PC-1D model [14], and can be expressed as

$$\frac{n}{n+p} \cdot J_p - \frac{p}{n+p} \cdot J_n = qR_{\text{surf}}. \quad (8)$$

The surface recombination rate, R_{surf} , given by the usual expressions [9], accounts for the influence of traps and dangling bonds at the semiconductor-metal interface. The additional constraints placed upon the system are

$$J_n + J_p = J_{\text{ext}} \quad (9)$$

$$\epsilon \frac{\partial \Psi}{\partial n} = Q_{\text{int}} \quad (10)$$

which simply states that the external current is equal to the sum of the carrier currents at the contact and that the gradient of the potential at the contact is proportional to any charge at the interface. Combining (8) and (9) yields expressions setting the distribution of currents into a contact

$$J_p = qR_{\text{surf}} + \frac{p}{n+p} \cdot J_{\text{ext}} \quad (11)$$

$$J_n = -qR_{\text{surf}} + \frac{n}{n+p} \cdot J_{\text{ext}}. \quad (12)$$

These current densities are used directly within the control volume formulation of the discretized continuity equations. A full two-dimensional (2-D) analysis of the diode archetypes used in this study is warranted if effects such as edge breakdown of the planar structure, gain nonuniformity either within a pixel or across pixels, or electrical crosstalk are of interest. However at this time, we are only interested in the charge storage related aspects of these devices and consequently it is possible to treat them as one-dimensional devices. This assumption allows the analysis to be independent of geometrical specifics, and most importantly, the external current density, J_{ext} , at the current specified boundary condition is now proportional to the total current by the cross-sectional area of the contact. For 2-D device simulation, this boundary can be utilized either by equipartitioning the current densities between nodes along the contact, imposing an additional constraint upon the sum of the currents flowing out of adjacent nodes comprising the contact, or by using a technique involving the additional simulation of the contacting metal [15].

The photo-illumination model used in these simulations is representative of the moonless, cloudless night sky. The input spectrum is taken from experimental measurements which include the contributions from airglow, stars, and zodiacal light over the range of 0.3–1.1 μm with the nominal illumination of 5.4×10^{16} photons per cm^2 per second (p/s) [16]. A standard wavelength-dependent absorption coefficient for silicon is also used [17]. The total thickness of both devices is set at 50 μm , yielding an internal quantum efficiency of approximately 55% for this spectrum. The lateral dimensions of the diodes conform to typical pixel geometries of 25 $\mu\text{m} \times 25 \mu\text{m}$.

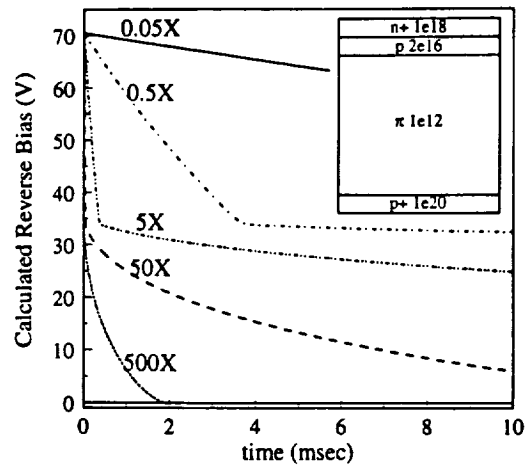


Fig. 1. Voltage recovery of reach through APD versus time for five magnitudes of incident photo-illumination. Inset illustrates salient geometry of RAPD device under examination.

III. RESULTS OF SIMULATION

A. APD Devices

The principal objective behind modeling APD devices in this manner is to obtain a general picture of their performance when used in a flux integrating mode. To this end, two basic generic devices are examined. The first is a reach-through APD (RAPD) as illustrated in the inset of Fig. 1 [18], [19]. This diode consists of a 2- μm n^+ cap layer above a 1.5- μm p-type multiplication layer. Below this is a large intrinsic layer and a thin highly-doped substrate making the entire structure 50 μm in length. By design during the initial reset, a large electric field exists in the p-type layer and a modest drift field is established in the intrinsic region. Because of the illumination spectrum and the geometry of the device, the majority of the carriers are generated in the large intrinsic drift region where the electrons must be transported to the high field multiplication layer before the signal may be amplified through impact ionization. For a biased diode these photo-excited carriers are swept across the depletion region, collected at the contacts, and the signal is detected through the fluctuation in the external current within the circuit. However, for the case at hand there is no direct connection to the power supply during the integration period. The generated carriers are still transported across the depletion region where they may or may not undergo impact ionization depending on the magnitude and extent of the electric field. Following that, however, the charge is stored near the edges of the high field region within the device and does not interact with the external circuit. Therefore as charge accumulates within the depleted diode, the effective bias across the device is decreased, and hence drives the reverse recovery.

The second design under consideration is a simple p-n junction APD as shown in Fig. 2. This device is very similar to the first except there is no intrinsic drift region and the p-type doping extends for the full length of the device. All carriers generated in the undepleted area must diffuse to the high field region before impact ionizing. Although the diffusion process leads to a slower response than the field aided within

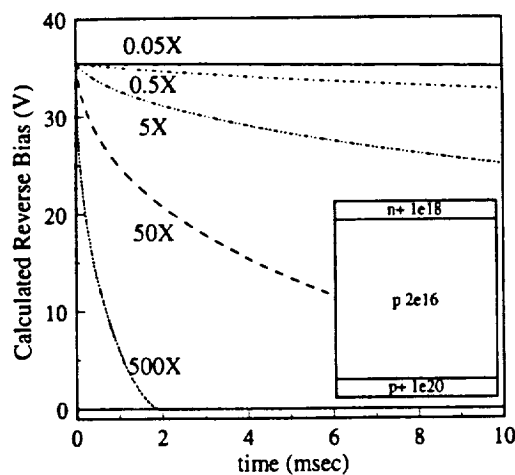


Fig. 2. Reverse recovery of p-n junction APD device. Internal bias is plotted against time for the five different incident photo-illumination powers under consideration. Inset shows overall p-n junction device geometry.

the RAPD, the temporal response of the device is not a key concern here due to the rather large charge collection times under examination. Furthermore, since only a portion of the diode is depleted, the initial operating bias of the device is considerably lower. However, the nature of the diffusion process is in general omni-directional, therefore pixel crosstalk may be a concern and additional design complexity will be required to minimize image blurring.

B. APD Performance

The model described above is used to study the two sample APD devices. At the beginning of the simulation, the diodes are biased to their steady-state operating voltage. The bias is then removed and the current controlled boundary condition applied. At this point, the recovery of the diodes is numerically calculated. Within this study the photoresponse, effects of variations within the bulk lifetime, and the sensitivity of performance to the initial diode gain are numerically examined and comparison is made between the two device types.

During the application of the reset bias, energy is stored within the internal depletion capacitance of the diode. Once the reset bias is removed and the cathode voltage freely floats, several mechanisms operate to neutralize the displaced charge either through an external leakage current or from generation of charge within the device itself. In the present study, the external leakage current from the read-out circuitry, J_{ext} , is neglected as is the surface recombination rate. The primary dark current source is the SRH mechanism due to the relative weakness of radiative and Auger recombination at the doping concentrations considered here. Except where specified, the SRH lifetimes for both electrons and holes are set at 100 μ s.

The first example we consider of the charge storage dynamics is for the case where photo-excited carriers are generated and collected within each device. Figs. 1 and 2 show the calculated reverse bias of the two diodes for illumination intensities ranging from 0.05 to 500 times the night sky spectrum (2.7×10^{15} – 2.7×10^{19} photons per cm^2 p/s). The considerable difference between the initial bias of the two

devices is due exclusively to the additional potential required to deplete the intrinsic region of the RAPD owing to the fact that the breakdown field in the multiplication region should be essentially the same between the two. Initially, both structures are depleted and the initial gain is set to 15.

For the RAPD, there can be up to three regimes of operation depending upon the quantity of photo-illumination. As carriers are generated within the device, they are swept to the edges of the depletion region and stored. Additionally, any secondary carriers generated from impact ionization are also transported to the edge of the high field region. The total amount of charge which can be stored in the intrinsic region is approximately equal to the total amount of charge depleted during reset. The holes collected in the RAPD structure initially drift into the depleted intrinsic region. Due to the low doping, this region will fill rather quickly resulting in a rapid decrease in the calculated bias. This effect is evidenced in the simulation and can be observed directly from Fig. 1. For very low light levels the intrinsic region is never filled and a linear decrease in the bias is observed. However, this is not seen for higher illumination since the intrinsic region is quickly filled by stored charge. To increase the amount of storage in the drift region, the background doping can be moderately increased at the expense of the linearity and required initial bias. Up to the point of filling the intrinsic region, the gain within the device is essentially constant for all light levels. The second regime of operation occurs once the drift region has been filled and charge is stored at the edge of the high field multiplication region resulting in the degradation of the gain within the device. This reduction in gain is again dependent on the rapidity of charge being added to the structure and for this device the internal gain is negligible when the bias drops below 10 V. In the extreme case of high illumination intensity, the diode is completely restored to its equilibrium condition and then actually begins to become forward biased. Recombination processes prevent the diode from continued storage of charge once the diode becomes sufficiently forward biased.

The reverse recovery rate of the p-n diode decays in much the same way as the RAPD with the exception of the initial recovery as seen in Fig. 2. Since the diode doesn't contain a low-doped drift region, the collected charge is immediately stored at the edge of the high field multiplication region. The total decay is quantitatively similar to the second regime of the RAPD. For the three lowest light levels, the voltage is essentially linear over the time frame of interest.

It is perhaps, more interesting to examine the actual stored charge within the device rather than the voltage drop resulting from its recovery. Typically, it is the stored charge which is sampled in the read-out circuitry [20]. Figs. 3 and 4 illustrate the total quantity of electron charge stored within each of the two devices as a function of integration time for the same variation in incident light levels as discussed above. During the first part of the integration, charge is collected in a linear fashion. As the total amount of charge increases, the gain begins to saturate resulting in an asymptotic square root dependence with time [8], [19]. However, this gain saturation is dependent upon the device being aptly biased. Once the diodes have recovered to the point where impact

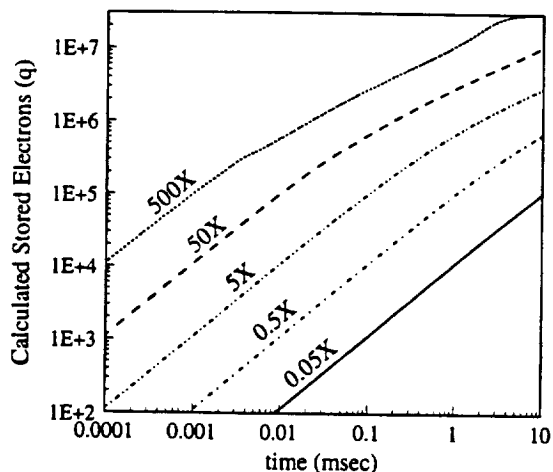


Fig. 3. Total stored charge within RAPD device during recovery. Incident illumination power corresponds to that in Fig. 1.

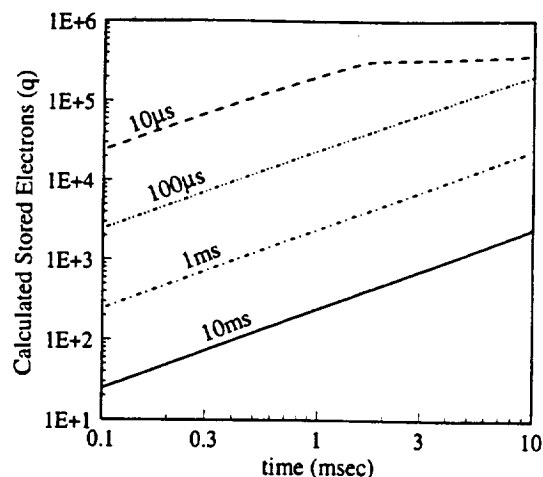


Fig. 5. Calculated stored charge in RAPD device for various SRH lifetimes.

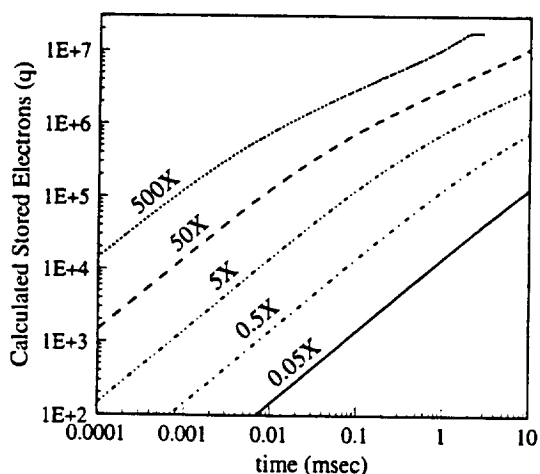


Fig. 4. Total stored charge in p-n junction device versus integration time.

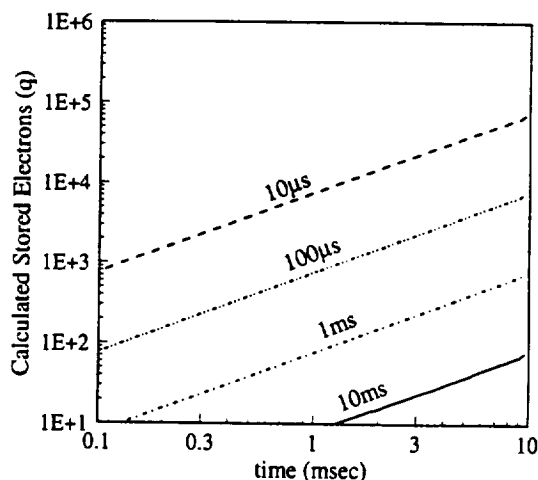


Fig. 6. Calculated stored charge in p-n junction device for various SRH lifetimes.

ionization is negligible, a trend toward a linear recovery is again observed. This result is clearly seen in the figures for the highest illumination level. Following the full voltage recovery, any additional charge will tend to forward bias the device analogously to an open-circuited solar cell. As seen in the figure, the stored charge in the p-n junction device saturates very abruptly as the additional carriers quickly drive the narrow junction into a high level injection state. However in the RAPD, carriers can easily spread into the low-doped region thus increasing the total capacity of the device.

The next issue explored in characterizing the performance of flux integrating diodes is to look at how the Shockley-Read-Hall lifetime parameters affect the performance of the charge integration. For a CCD, it is absolutely critical that all defects be removed from the device if the transfer efficiency is to be near unity. Since the APD charge storage devices will be operated as pixels, it may be possible that lower grade material can be used than is necessary for CCD's. Simulations are performed with various generation lifetimes ranging from 10 ms to 10 μs [21]. Figs. 5 and 6 show the results of these simulations for the two devices under dark conditions where thermal generation is the sole source

of additional carriers. As expected, the RAPD has a much higher rate of recovery due to the larger initial volume of the depletion region. As seen in Fig. 5, once the intrinsic region in the RAPD has been filled, the two devices recover at the same rate as exemplified in the 10-μs lifetime simulation. An interesting result of this study is that even the lowest photoillumination rate, which is well below the clear night sky spectrum, produces a comparable amount of charge to that of the worst lifetime simulation which is a full three orders of magnitude lower than what is technologically achievable. This is a very promising result indicating that low-cost silicon technology may be applicable for room-temperature low light level sensing in this spectral range. Other narrow-gap material systems which would benefit from greater spectral response would most likely require expensive cooling to reduce the dark current, thus prohibiting its use in low cost applications.

A limiting factor in the application of APD arrays for low light level imaging is their lack of gain uniformity both within individual pixels or between various pixels in the array. Planar APD arrays suffer from pixel gain nonuniformity introduced either from processing variations within the devices, the edge breakdown suppression mechanism, or the dead regions arising

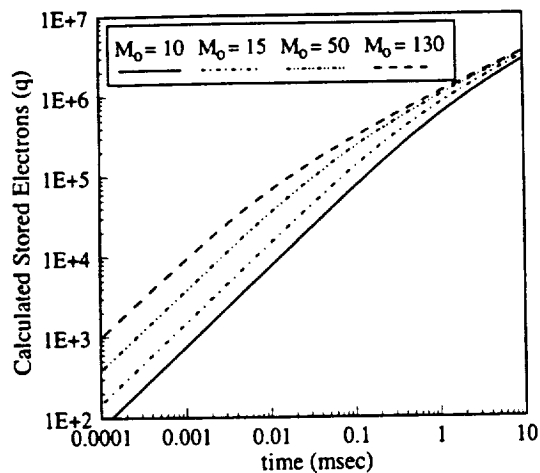


Fig. 7. Total stored charge in p-n junction device versus integration period. The four curves correspond to the calculated response for different initial reset biases. The initial gain at the reset bias is given. The illumination power used for these simulations is five times the night sky spectrum.

from device isolation. However, techniques which rely upon etching and grooves to isolate the devices have been demonstrated to possess good uniformity [22]. Also, the combined effects of saturation in high gain regions and the integration of flux over a relatively long time frame will tend to reduce the magnitude of the nonuniformity within a pixel.

Clearly, processing variations will limit the amount of uniformity between adjacent devices in an array, and subsequently will introduce a spread in the breakdown voltages between them. To examine the effects of nonuniformity between devices, this study also looks at the diode performance as a function of the initial bias. Fig. 7 illustrates the results of simulating the p-n diode under illumination for different initial bias and, hence, gain. As seen in the figure, during the beginning of the integration period, the devices behave linearly with the difference in stored charge being proportional to the difference in gain. As before, the stored charge begins to saturate as time increases. The time to the onset of saturation is inversely proportional to the square of the initial gain which is again consistent with the classical theory of biased APD's [19]. From the figure, it is observed that for moderate to high light levels, the importance of processing nonuniformities contributing to variations in the initial gain are reduced because of the self-quenching mechanisms inherent in the devices.

IV. CONCLUSIONS

The design of an APD array for use as an image sensor in the charge storage mode depends critically upon the structure of the basic diode prototype. In this study, two distinct silicon APD device designs are examined and an analysis of the operation of these devices in the charge storage mode is presented. In order to accurately model such diodes and to account for nonlinearities in their operation, an extensive numerical analysis is necessary. To properly model the recovery response using a numerical simulator, a new boundary condition is introduced. This new condition allows the external current of the device to be the specified, independent quantity, rather

than the bias. With this, the voltage recovery of the floating contact of the diode may be directly observed without relying upon an arbitrary external circuit. Using the presented model with the new boundary condition, reach through and a p-n junction APD's are examined specifically to observe the diode recovery, the total stored charge, the influence of the recombination lifetimes, and the effect of initial device gain upon the transient response.

The two device designs recover in the same manner with a few notable exceptions. The RAPD requires a significantly higher reset bias in order to deplete the entire drift region. This drift region not only allows the carriers to move rapidly to the multiplication region, but the design will be advantageous in terms of pixel cross-talk whereas the standard p-n junction device may be prone to image blur unless additional isolation is employed between pixels. Furthermore, clever design of the RAPD device may allow for edge breakdown suppression [23]. In terms of charge storage, both devices show evidence of gain saturation during recovery due to the degradation of the internal bias resulting in a square root dependence of stored charge upon the integration time. As a large fraction of the bias is recovered, the carrier multiplication is extinguished, and both diodes begin to store charge at a linear response rate, albeit at unity gain. The maximum storage capacity of the devices is again dependent upon the device geometry. Since the RAPD requires a greater depletion of charge during reset, it has a larger capacity. Also, once the diodes have fully recovered, any additional accumulated charge acts to forward bias the devices. The overall capacity limit is set by the recombination processes and the distribution of charge within the forward biased devices. The RAPD will clearly go into a high level injection state far after the p-n junction device since the excess charge can easily be redistributed within the large, low-doped region. Thus, the RAPD device has a substantially larger total capacity. The dark response of the two devices is examined for recombination lifetimes between 10 ms and 10 μ s. The recoveries of the two devices here are considerably different due to the vast difference between the volume of their depleted regions. For the RAPD, once the intrinsic drift region fills with charge, it is observed that the recovery continues at the same rate as the p-n junction device as expected.

Finally, the effects of initial gain are observed to impact the transient response of the device. Probably the largest limitation in the construction of APD arrays for low light level imaging arises from the lack of uniformity between pixels. Based on this, large scale arrays would be difficult to achieve. However, it is observed in this study that the sub-linear collection of the charge for large initial gains acts to cancel any discrepancy in accumulation caused by the dissimilarity of devices. Therefore, the effects of charge saturation when the devices are used in a flux integration mode, rather than using direct readout, will help reduce some of the problems associated with device nonuniformity in an APD array.

REFERENCES

- [1] T. Ando, M. Yasuda, and K. Sawada, "Noise suppression effect in an avalanche multiplication photodiode operating in a charge accumulation

- mode," *IEEE Trans. Electron Devices*, vol. 42, pp. 1769–1774, Oct. 1995.
- [2] H. Komobuchi and T. Ando, "A novel high-gain image sensor cell based on Si p-n APD in charge storage mode operation," *IEEE Trans. Electron Devices*, vol. 37, pp. 1861–1868, Aug. 1990.
 - [3] H. Komobuchi, M. Morimoto, and T. Ando, "Operation and properties of a p-n avalanche photodiode in a charge integrating mode," *IEEE Electron Device Lett.*, vol. 10, pp. 189–191, May 1989.
 - [4] E. Fossum, "Active pixel sensors: Are CCD's dinosaurs?," *SPIE*, vol. 1900, pp. 2–14, 1993.
 - [5] T. P. Lester and D. L. Pulfrey, "A new MOS photon-counting sensor operating in the above-breakdown regime," *IEEE Trans. Electron Devices*, vol. ED-31, pp. 1420–1427, Oct. 1984.
 - [6] G. Weckler, "Operation of p-n junction photodetectors in a photon flux integrating mode," *IEEE J. Solid-State Circuits*, vol. SC-2, pp. 65–73, Sept. 1967.
 - [7] A. W. Smith, "Light confinement and hydrodynamic modeling of semiconductor structures using volumetric techniques," Ph.D. dissertation, Georgia Institute of Technology, Atlanta, 1992.
 - [8] J. W. Parks, A. W. Smith, K. F. Brennan, and L. E. Tarof, "Theoretical study of device sensitivity and gain saturation of separate absorption, grading, charge, and multiplication InP/InGaAs avalanche photodiodes," *IEEE Trans. Electron Devices*, vol. 43, pp. 2113–2121, Dec. 1996.
 - [9] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. New York: Springer-Verlag, 1984.
 - [10] S. V. Pantankar, *Numerical Heat Transfer and Fluid Flow*. New York: Hemisphere Publishing Corp., 1980.
 - [11] T. V. Oppé, W. D. Joubert, and D. T. Kincard, "An overview of NSPCG: A nonsymmetric preconditioned conjugate gradient package," *Computer Physics Comm.*, vol. 53, pp. 283–292, 1989.
 - [12] W. Kausel, J. O. Nylander, G. Nanz, S. Selberherr, and H. Poetzel, "BAMBI: A transient 2D-MESFET model with general boundary conditions including Schottky and current controlled contacts," *Microelectron. J.*, vol. 21, no. 5, pp. 5–21, 1990.
 - [13] N. Iwamuro and S. Tagami, "Stable calculation for power device simulation with inductive load circuit: Application of an integral method," *Microelectron. J.*, vol. 24, pp. 139–145, 1993.
 - [14] P. A. Basore and D. T. Rover, *PC-ID Installation Manual and User's Guide*, Iowa State Univ. Res. Foundation, Ames, 1985.
 - [15] P. A. Gough, M. K. Johnson, S. A. Higgins, J. A. G. Slatter, and K. R. Whight, "Two-dimensional simulations of power devices with circuit boundary conditions," *NASECODE V*, 1987, pp. 213–8, 1987.
 - [16] R. W. Engstrom and R. L. Rodgers, "Camera tubes for night vision," *Opt. Spectra*, vol. 26–30, pp. 438–442, Feb. 1971.
 - [17] H. Melchior, "Demodulation and photodetection techniques," in *Laser Handbook*, F. T. Arecchi and E. O. Schultz-Dubois, Eds. Amsterdam, The Netherlands: North Holland, vol. 1, pp. 732, 1972.
 - [18] R. J. McIntyre, "Recent developments in silicon avalanche photodiodes," *Measurement*, vol. 3, no. 4, pp. 146–152, 1985.
 - [19] P. P. Webb, R. J. McIntyre, and J. Conradi, "Properties of avalanche photodiodes," *RCA Rev.*, vol. 35, pp. 234–278, 1974.
 - [20] S. Mendis, S. E. Kemeny, and E. R. Fossum, "CMOS active pixel image sensor," *IEEE Trans. Electron Devices*, vol. 41, pp. 452–453, Mar. 1994.
 - [21] P. T. Landsberg, "Trap-auger recombination in silicon of low carrier densities," *Appl. Phys. Lett.*, vol. 50, no. 12, pp. 745–747, Mar. 1987.
 - [22] M. R. Squillante, J. S. Gordon, R. Farrell, S. Vasile, K. Delay, C. Oakes, and K. Vanderpuye, "Recent advances in avalanche photodiode technology," *SPIE—X-Ray Detector Physics and Applications II*, vol. 2009, pp. 64–71, 1993.
 - [23] P. P. Webb and R. J. McIntyre, "Multi-element reachthrough avalanche photodiodes," *IEEE Trans. Electron Devices*, vol. ED-31, pp. 1206–1212, Sept. 1984.

Joseph W. Parks, Jr., photograph and biography not available at the time of publication.

Kevin F. Brennan, photograph and biography not available at the time of publication.

A Smoothed Boundary Condition for Reducing Nonphysical Field Effects

Arlynn W. Smith, Joseph W. Parks, Jr.,
Joe N. Haralson II, and Kevin F. Brennan

Abstract—In this paper, we examine the problem associated with abruptly mixing boundary conditions in the context of a two-dimensional semiconductor device simulator. Explicitly, this paper addresses the transition between an ohmic-type Dirichlet condition and a passivated Neumann boundary. In the traditional setting, the details of the transition between the two boundary types are not addressed and an abrupt transition is assumed. Subsequently, the calculated observables (most notably the potential) exhibit discontinuous derivatives near the surface at the point where the boundary type switches. This paper proposes an alternative condition which models the progression between the two boundary types through the use of a finite length, smoothed boundary whereby the numerical discontinuities are eliminated. The physical and mathematical basis for this smoothed boundary condition is discussed and examples of the technique's implementation given. It is found that the proposed boundary condition is numerically efficient and can be implemented in pre-existing device simulators with relative ease.

Index Terms—Boundary conditions, macroscopic device simulation.

In the two-dimensional (2-D) simulation of semiconductor devices such as metal–semiconductor–metal (MSM) photodetectors [1]–[3], MOSFET's, etc., there is often a switching of boundary conditions along the lateral surface of the device between a metal contact and a passivated region. Due to the nature of how these boundaries are modeled, the mathematics forces a change from a Dirichlet- to a Neumann-type boundary condition somewhere along the edge of the device. Application of a mixed Dirichlet and Neumann boundary condition could, in principle, rectify this problem. However, an unambiguous specification of this transition throughout the contact surface introduces an additional constraint. To our knowledge, specification of this additional constraint cannot be simply derived or formulated with much accuracy. Typically, the change in boundary condition is treated as an ideal, abrupt transition for lack of a better description and further knowledge of the nature of the interface. The calculated observables, therefore, can exhibit discontinuous derivatives laterally along the boundary. This poses a multifaceted problem based on both physical and mathematical grounds. For instance, as the applied bias on the conduction contact is increased, the discontinuity in the boundary condition leads to the prediction of very high, nonphysically based electric field spikes near the interface discontinuity. The erroneously calculated electric field magnitude can then dramatically affect such parameters as the impact ionization and tunneling rates and the mobility resulting in the possibly incorrect estimation of carrier transport and breakdown effects. Continued mesh refinement only exacerbates the problem by making the finite difference approximations to the derivatives

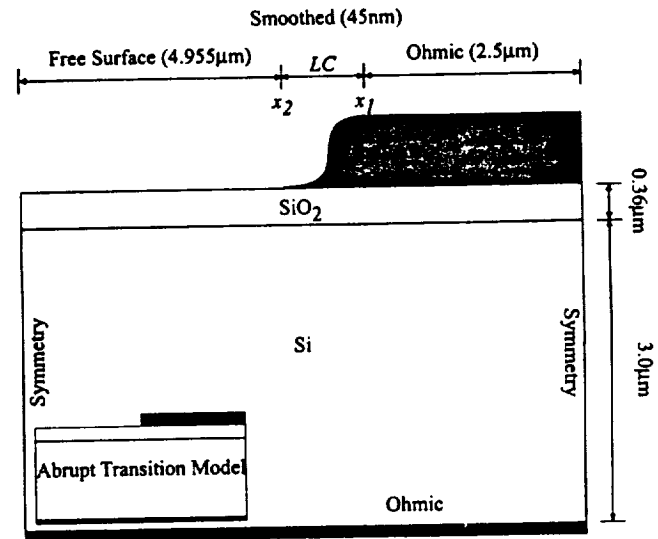


Fig. 1. Illustration of sample device using the new smooth boundary condition under consideration. The smoothed contact regime is along the top surface of the device between x_1 and x_2 . The inset schematically shows what is considered as an abrupt boundary transition.

more singular. Additionally, output current characteristics are often desired at the metal contact. The presence of these discontinuous derivatives makes the calculation of the current at the edge of the metal difficult [4], [5]. From a mathematical point of view, the Laplacian operator in Poisson's equation requires that the potential be at least C^1 continuous within the simulation domain and along the domain boundary. Using the abrupt interface between boundary conditions does not preserve the required continuity, and thus fails to guarantee the correct solution of the equation. This is exemplified numerically through unbounded or oscillating solutions whereby the system never reaches a converged result. One very simplistic means of reducing the unwanted discontinuities is to use a coarse mesh in the vicinity of the boundary change. Doing that, however, can remove the possibility of accurately capturing the true physics of the device such as current crowding effects.

Alternatively, a physically accurate treatment of the contact metallization is impractical. A pictorial example of the difference between an abrupt and a more physically realizable metallization pattern is given in Fig. 1. An accurate incorporation of the "tailing" of the metal is difficult to implement. First, the conductivity of the metal as a function of the distance from the edge of the contact and applied bias is generally not known. Moreover, and more importantly, the approximation of an equipotential surface breaks down as the metal is thinned. As the work function of the metal is used to pin the Fermi level of the underlying material, any deviations from an equipotential surface in the metal leads to a violation of the assumptions used in the derivation of an ideal ohmic contact. Simulators which specify a finite resistivity for the metal, but then fix the potential and carrier concentration to values based on an ideal ohmic contact [6], only address current flow to an external circuit, not the problem of an abrupt change in the boundary condition. It is this last concern that we wish to address.

In the modeling of devices, it is important to construct an approach which accurately determines the magnitude of the electric field along the lateral direction in the near vicinity of a metal–semiconductor or metal–insulator interface. This technique should remove any

Manuscript received December 1, 1995; revised February 13, 1996, April 27, 1996, and February 13, 1997. This work was supported in part by ARPA through the Phosphor Technology Center of Excellence under Contract MDA972-9301-0030 and by ARPA through contract to NASA. NAGW-2753. This paper was recommended by Associate Editor S. Duvall.

A. W. Smith was with the Georgia Institute of Technology, Atlanta, GA 30332-0250 USA. He is now with ITT Electro-Optical Products Division, Roanoke, VA USA.

J. W. Parks, Jr., J. N. Haralson, II, and K. F. Brennan are with the School of Electrical and Computer Engineering, and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

Publisher Item Identifier S 0278-0070(97)05148-8.

numerical artifacts produced through the discretization of the model equations without severely changing the overall characteristics of the device. In order to improve the predictive capability of numerical device simulators for structures which contain an abrupt boundary condition, we propose a new technique for smoothing the boundary condition based on both physical and mathematical arguments. Rather than having an abrupt switch in boundary condition, a short transition, or "smoothed," boundary is employed. The only premise behind this contact type lies in ensuring the conservation of the required continuity along the periphery of the simulation domain. The parameters used for the calculated observables on either side of the smoothed contact are given by the traditional boundary types. The smoothed contact requires the continuity of the potential and its first derivative along the lateral direction at either edge of the boundary to be maintained. With respect to Fig. 1, the following four conditions are mandated:

$$\begin{aligned}\Psi(x_1) &= \Psi_1 & \Psi'(x_1) &= \Psi'_1 \\ \Psi(x_2) &= \Psi_2 & \Psi'(x_2) &= \Psi'_2.\end{aligned}\quad (1)$$

The simplest functional dependence which forces this condition is a cubic polynomial. Substituting the requirements of (1) into the cubic yields an expression for the smoothed potential along the surface of the device

$$\begin{aligned}\Psi(x) &= \left[\frac{2(\Psi_1 - \Psi_2) + \Psi'_1 LC + \Psi'_2 LC}{LC^3} \right] (x - x_1)^3 \\ &+ \left[\frac{\Psi'_2 - \Psi'_1}{2LC} + \frac{6(\Psi_2 - \Psi_1) - 3\Psi'_1 LC - 3\Psi'_2 LC}{2LC^2} \right] \\ &\cdot (x - x_1)^2 + \Psi'_1 (x - x_1) + \Psi_1\end{aligned}\quad (2)$$

where $LC = x_2 - x_1$ is the length of the smoothed boundary. The values of Ψ_1 and Ψ_2 are obtained from the solution of the previous iteration, while Ψ'_1 and Ψ'_2 are first order finite difference approximations to the derivatives using the first two nodes outside the smooth regime. For the case presently under examination, the conditions on the metal side of the contact reduce to the applied bias and zero for Ψ and Ψ' , respectively. This new boundary type is treated as a Dirichlet contact and is implemented similarly to a metal contact. The overall system of equations is solved self-consistently over the entire simulation domain using this smoothed boundary, thereby assuring that the values for the potential are accurately calculated.

The simplicity of this technique lies in the fact that the only requirement on the boundary is the maintenance of the continuity of the potential. The aforementioned spikes arise from the inaccurate description of the boundary transition which is physically continuous yet often treated in a discontinuous manner. An increase in the magnitude of the electric field at the boundary interface is expected on physical grounds, but the standard treatment of the interface introduces an additional, purely numerical, spike at the discontinuity. The introduction of the smoothed boundary immediately introduces the required continuity, thus eliminating the numerically generated spike. This is guaranteed regardless of the length of the smoothed contact, provided that the domain discretization is sufficiently dense to accurately capture the nature of the transition. Additionally, this technique, unlike the abrupt case, is stable in the sense that continued mesh refinement about the transition will not greatly alter the results of the simulation.

Before simulations can be made utilizing this new boundary type, the additional degree of freedom embodied in the adjustable parameter LC needs to be specified. The size of the smoothed contact influences the distribution of the potential in the vicinity of the interface, and hence will affect the magnitude of the electric field in that region. A device with a very short transition length will develop a rather large, physically accurate, field near the interface. Conversely,

a device with a large transition length will have a lower overall field magnitude. The proposed smoothed boundary model predicts this trend in the field behavior, whereas the standard practice of using a discontinuous boundary does not. Under ideal conditions, an empirically based estimate for the "tail" length would be directly used as the smoothed boundary length in the calculation, requiring direct experimental measurement of LC . This would yield the best results for the most detailed analysis. However, recognizing that an exact value of this parameter is often unavailable, we have considered using a value of LC which is some fraction of the metal thickness of the ohmic contact. In this case, the actual numerical value is somewhat arbitrary, but it is seen that a modest range of values for LC exists such that the erroneous spikes are removed without greatly undermining the true field present within the device.

Fig. 1 shows a test device used to compare the effects of a smoothed boundary condition to that of an abrupt change in boundary type. The physical dimensions of the device are given in the figure, and the thickness of the ohmic metal contact is assumed to be 4500 Å. Fig. 2 displays the calculated potential along the lateral direction at the top of the device for the abrupt condition and for several values of LC as a percentage of the metal thickness. From the figure it is clearly seen that employing the smoothed boundary condition has not appreciably altered the overall potential profile of the device. However, near the interface where the smoothed boundary condition has been used, a difference is observed. The inset of Fig. 2 illustrates the potential in the immediate neighborhood of the abrupt discontinuity. For the case of the abrupt boundary transition, the slope of the potential is discontinuous while the smoothed boundary condition obeys the requirements which were imposed by the cubic polynomial given by (1) and (2).

Figs. 3 and 4 show the calculated magnitude and vector direction of the electric field at a bias of 12 V for both the abrupt transition and a case where a 10% smoothed boundary is used. Only a very small portion of the device is shown to emphasize the boundary transition. Looking at the overall device, it is observed that the field spike is confined to a very small region near the surface, and that the field even one tenth of a micron away from the contact is essentially unchanged. For the MIS structure under consideration, the field discontinuity is on the oxide surface and, subsequently, both the abrupt and the smoothed cases give roughly the same field inside the semiconductor. The large spike in the abrupt case clearly originates from the discontinuity. In the smoothed case, the spike has been removed yielding a more physically realistic electric field at the interface. The vector plots illustrate that the direction of the field changes in a consistent and continuous manner using the smoothed boundary condition. When the abrupt boundary condition is employed, the field direction undergoes a very sudden and dramatic shift in direction near the metal contact as can be seen from Fig. 3. In terms of convergence, the smoothed boundary, as a direct result of removing the numerical discontinuity, enables device simulation up to much higher biases. For example, the device with an abrupt boundary condition given in Fig. 3 did not yield a converged solution for biases greater than 12 V, while the device with the smoothed boundary condition was simulated to biases exceeding 60 V. The ability to simulate high voltage conditions is often very important as, for example, in the case of avalanche photodiodes.

By performing several simulations, it is further observed that additional mesh refinement directly, and undesirably, impacts the magnitude of the field for the abrupt case while the smoothed condition is insensitive to additional refinement. To illustrate the stability of this simulation with the new boundary condition, the maximum electric field as a function of grid refinement is plotted in Fig. 5 at a constant value of LC . As can be seen from inspection of Fig. 5, when the new boundary condition is employed, the

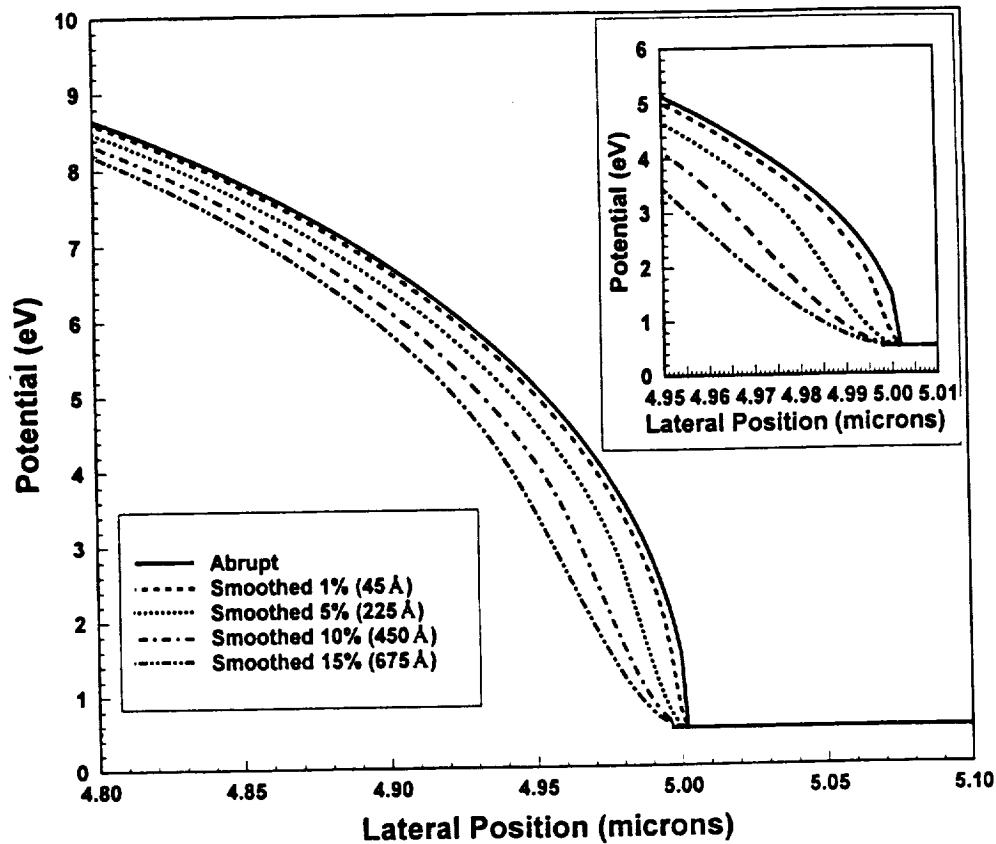


Fig. 2. Electrostatic potential along the top surface of the simulation domain in the region of the boundary transition. The potential for several values of LC in addition to the abrupt case are given. The inset shows an enlarged portion of the graph in the immediate vicinity of the transition.

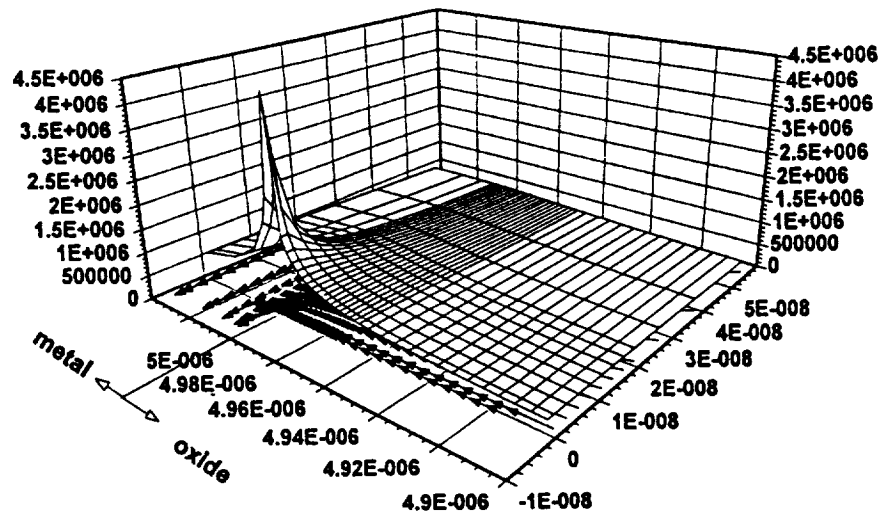


Fig. 3. Magnitude and vector plot of the electric field for the abrupt transition near the top surface of the device.

maximum electric field remains essentially unchanged with grid refinement. Alternatively, if the standard approach is employed, the maximum electric field greatly increases with increasing grid refinement. As discussed above, the difference in the maximum electric field between the two methods is due to the mathematical error introduced in the abrupt formulation. In the limit of a very coarse mesh, both techniques converge to the same maximum electric field. Therefore, the application of the new boundary condition results in an unambiguous specification of the magnitude of the electric field independent of the degree of mesh refinement.

As previously mentioned, the distance over which the boundary is smoothed, LC , is expected to alter the final results of the simulation. This is shown graphically in Fig. 6 for the device under consideration. Here, the maximum electric field is plotted against the length of the smoothed boundary at a bias of 4 V. For the range of LC considered here, the electric field varies by about a factor of three with an observed saturation for large values. This range is due to the difference between the physically realistic potential profiles and is not a consequence of numerical error as seen in the abrupt case. Even if the length of the smoothed boundary is completely

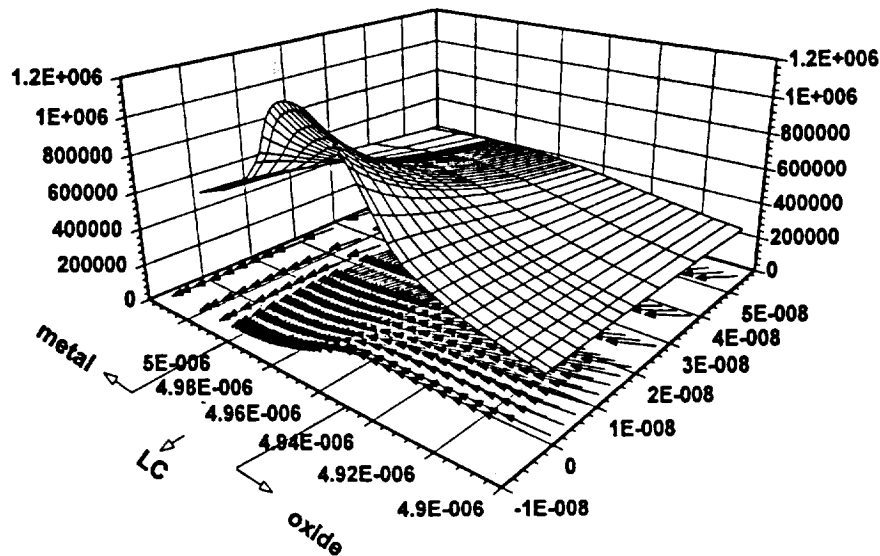


Fig. 4. Same as Fig. 3 using the results of a 10% smoothed boundary region ($LC = 450 \text{ \AA}$).

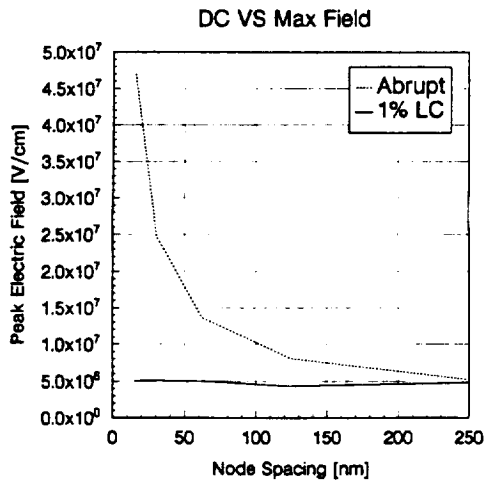


Fig. 5. Illustration of the maximum electric field versus the uniform distance between nodes for the abrupt and smoothed boundary conditions. For the smoothed case, LC is constant at 1% of the metal thickness. Applied bias is 12 V for all cases.

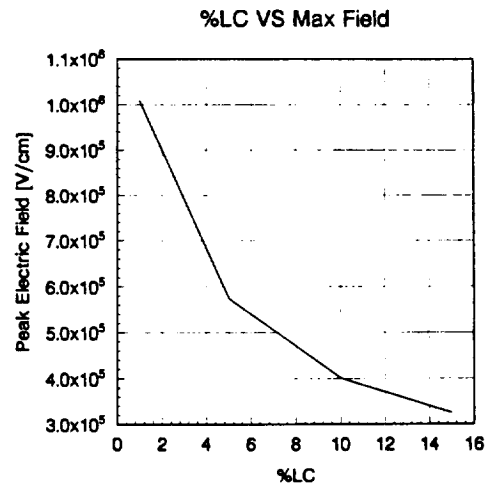


Fig. 6. Maximum electric field calculated for different values of LC at an applied bias of 4 V.

unknown, the maximum uncertainty in the estimation of the field is still considerably smaller than that associated with using the abrupt boundary condition. The smoothed boundary, therefore, is again seen to be an improvement over the standard discontinuous model.

We have presented a new approach for modeling a change in the boundary condition along the lateral direction of a metal-semiconductor or metal-insulator interface. The technique introduces a small interface region between the two incompatible boundary types which is designed to ensure the necessary continuity of the potential and its first derivative along the lateral direction at both ends of the smoothed boundary. Using this new technique, it is observed that the electric field at the interface becomes insensitive to the degree of mesh refinement, as opposed to the case of an abrupt transition where the magnitude varies strongly with the mesh. Therefore, the new method leads to an unambiguous specification of the magnitude of the electric field, resulting in a more accurate and physically realistic description of the device performance. Just as the values of the device geometry and doping profiles must be known prior to conducting a simulation of a device, the additional parameter LC must be specified. As is often the case with other

geometrical features of a device, the exact value of LC may not be known to within any great accuracy as it corresponds to the length of the metal "tail" of the conducting contact. It is observed, however, for the values of LC considered in this paper that the "smoothed" boundary does not substantially affect the magnitude of the electric field within the semiconductor.

REFERENCES

- [1] L. Figueroa and C. S. Slayman, "A novel heterostructure interdigitated photodetector (HIP) with picosecond optical response," *IEEE Electron Device Lett.*, vol. EDL-2, pp. 208-210, Aug. 1981.
- [2] S. Y. Chou and M. Y. Liu, "Nanoscale tera-hertz metal-semiconductor-metal photodetectors," *IEEE J. Quantum Electron.*, vol. 28, pp. 2358-2368, Oct. 1992.
- [3] A. F. Salem and K. F. Brennan, "Theoretical study of the response of the InGaAs metal-semiconductor-metal photodetectors," *IEEE J. Quantum Electron.*, vol. 31, pp. 944-953, May 1995.
- [4] E. Palm and F. Van De Wiele, "Current lines and accurate contact current evaluation in 2-D numerical simulation of semiconductor devices," *IEEE Trans. Electron Devices*, vol. ED-32, pp. 2052-2059, Oct. 1985.
- [5] G. Nanz, P. Dickinger, and S. Selberherr, "Calculation of contact currents in device simulation," *IEEE Computer-Aided Design*, vol. 11, pp. 126-136, Jan. 1992.
- [6] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. New York: Springer-Verlag, 1985.

Theoretical Calculations of Charge Confinement in a $pn^{-}np$ Heterojunction Acoustic Charge Transport Device

Arlynn W. Smith, J. Stevenson Kenney, William D. Hunt, Kevin F. Brennan,
Senior Member, IEEE, Rudy Benz, and Christopher J. Summers, Member, IEEE

Abstract— An alternative structure for heterojunction acoustic charge transport (HACT) devices has been devised and analyzed. The structure uses a $pn^{-}np$ doping profile near the surface of the device to create a charge transport layer and provide top vertical confinement. This is contrary to previous n-type HACT structures which rely on residual surface states and a heterojunction discontinuity for the same functions. The use of the $pn^{-}np$ doping as the channel depletion mechanism makes the device insensitive to the residual surface state density, thus providing a more robust design. In addition, the use of the back np junction enables widening of the transport layer thereby increasing the amount of charge that can be transported by the acoustic wave. As a result of the increased charge capacity it is expected that the $pn^{-}np$ ACT device will exhibit a greater dynamic range and current than previous HACT designs. The analysis of the device structure is accomplished herein using a two dimensional hydrodynamic simulation code, Semiconductor Total Energy Balance Simulator in two Dimensions (STEBS-2D), which has been modified to account for the potential created by the surface acoustic wave. The calculated results indicate that an order of magnitude enhancement in charge capacity is possible using the new structure. Transfer efficiency calculations for several different lifetimes in the transport layer show high efficiency values, greater than 79% efficient with a Shockley Read Hall lifetime of 10 nsec.

NOMENCLATURE

ξ_r	Relative permittivity of semiconductor.
ξ_0	Permittivity of free space.
q	Electronic charge.
ψ	Electrostatic potential.
p	Hole concentration.
n	Electron concentration.
N_d^+	Ionized donor concentration.
N_a^-	Ionized acceptor concentration.
Γ_{SAW}	Effective doping due to surface acoustic wave.
t	Time.
$\nu_{n,p}$	Velocity of electrons and holes.
R,G	Recombination and Generation.

$\mu_{n,p}$	Electron and hole mobility.
k_b	Boltzmann constant.
$T_{e,h}$	Electron or hole temperatures.
ϵ_c	Conduction band edge (includes bandgap narrowing).
ϵ_v	Valance band edge (includes bandgap narrowing).
χ	Electron affinity.
E_g	Semiconductor band gap.
$m_{e,h}^*$	Electron and hole effective mass.
Λ_c	Ratio of Fermi integrals $\{\mathcal{F}_{1/2}(\eta_c)/\mathcal{F}_{-1/2}(\eta_c)\}$.
η_c	Reduced energy of the conduction band.
Λ_v	Ratio of Fermi integrals $\{\mathcal{F}_{1/2}(\eta_v)/\mathcal{F}_{-1/2}(\eta_v)\}$.
η_v	Reduced energy of the valance band.

I. INTRODUCTION

HETEROJUNCTION acoustic charge transport (HACT) devices are promising candidates for analog signal processors and rf memories. The HACT device structure is just one of the two major forms of acoustic charge transport (ACT) devices, bulk and surface types. In bulk ACT structures charge is confined in the vertical direction (~ 5 to $6 \mu\text{m}$ deep) using a bias potential created between a reverse-biased Schottky barrier and an applied back-gating potential. In HACT devices vertical confinement ($< 1 \mu\text{m}$ deep) is provided by discontinuities in the conduction band edge due to material changes. In both HACT and ACT devices lateral confinement and charge transport is accomplished using the potential created by a surface acoustic wave generated within the semiconductor. A schematic diagram of acoustic charge transport in a representative HACT structure is shown in Fig. 1. As can be seen from Fig. 1, the surface acoustic wave generated by a transducer on the surface of the semiconductor propagates through the structure pushing charge from the input contact over some distance for later extraction at an output contact.

The HACT structure offers several important advantages over a conventional buried channel ACT. The HACT device has less power dissipation, utilizes simpler channel isolation techniques, is self-biasing, and enables possible direct integration with other planar technologies [1]. These advantages are realized due to the shallower depth of confinement for the charge packet. Less input power is required because more efficient use of the surface acoustic wave (SAW) potential

Manuscript received May 2, 1994; revised December 21, 1994.

A. W. Smith, J. S. Kenney, W. D. Hunt, and K. F. Brennan are with the School of Electrical and Computer Engineering and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

R. Benz and C. J. Summers are with the Quantum Microstructures Branch, Georgia Tech Research Institute, Georgia Institute of Technology, Atlanta, GA 30332-0800 USA.

IEEE Log Number 9410170.

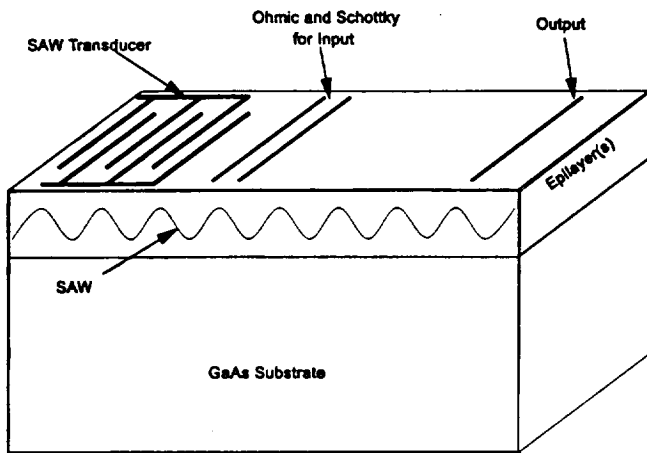


Fig. 1. Schematic diagram of a surface acoustic wave propagating in an epilayer structure.

is obtained from the shallower charge confinement. Channel isolation can be provided by mesa etching or proton isolation which would not be possible at a depth of $5\text{ }\mu\text{m}$. However, these advantages do not come without some penalty. In the conventional HACT (n-HACT) structure, shown in Fig. 2, the self-biasing of the device relies on residual surface states and very critical control of the doping in the layer structure. The one dimensional conduction band diagram, without any acoustic potential, for three different cases of residual surface charge for the n-HACT structure is displayed in Fig. 3, the surface charge densities are $-1.0 \times 10^{12}\text{ cm}^{-2}$, $-1.5 \times 10^{12}\text{ cm}^{-2}$, and $-5.0 \times 10^{12}\text{ cm}^{-2}$. The fermi level is at zero volts. Notice that at the lowest residual surface concentrations the channel is modulation doped by the dopants in the top AlGaAs layer. At moderate values of surface charge [2] the channel is depleted as required. This is the first surface state density for which we obtained charge confinement. As the surface charge increases the channel becomes more depleted. The depth of the depletion region is controlled by the density of the surface states and the doping concentration in the layers below. The design goal is to have the doping concentration in the AlGaAs layer slightly larger than the fixed surface charge [1]. Once the surface states have been satisfied this will then provide a small amount of mobile charge for transport [1]. A major problem is that the resulting value of surface charge is not known *a priori* i.e., different processing conditions or even different atmospheric conditions could change the final residual surface concentration.

Small variations in the doping concentrations within the layers can also upset the delicate charge balancing with the surface state density. Any of these deviations from perfect balance can cause large variations in the device performance. Relying on surface charge to deplete the channel layer also introduces a limit on the width of the transport layer, the surface charge acts over a small distance unless the material is lightly doped. A smaller transport layer thickness limits the charge capacity of the acoustic wave packet thereby decreasing the dynamic range of the device. Therefore, a need exists for a HACT device design which is insensitive to the residual surface states and has performance characteristics comparable

20 nm	GaAs	N-type NID
70 nm	AlGaAs with 30% Aluminum	N-type 2.0×10^{17}
40 nm	GaAs	N-type NID
100 nm	AlGaAs with 30% Aluminum	N-type NID
GaAs Substrate		N-type NID

Fig. 2. Epitaxial layer structure of a conventional n-type heterojunction acoustic charge transport device (n-HACT).

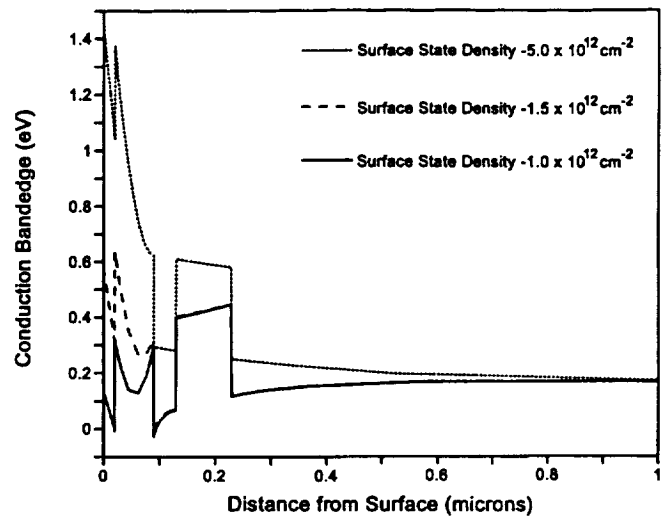


Fig. 3. One dimensional conduction band diagrams of the n-type HACT device with three different residual surface state concentrations, no acoustic potential.

to or greater than the current HACT structures which rely on the surface states and have limited channel widths. A structure which utilizes a pn junction at the surface to isolate the channel from surface states was proposed in reference [3]. However, the structure was never analyzed to determine the necessary doping profiles or the device characteristics. The structure proposed in this paper is an extension of the pn HACT configuration first described in reference [3].

The purpose of this paper is to introduce and analyze an alternative HACT structure, hereafter called p-HACT, which is relatively insensitive to changes in residual surface concentration and to fluctuations in the doping concentration of the epitaxial layers. The proposed device structure utilizes a pn junction at the front surface to provide isolation from the surface charge and to create the top vertical confinement. A lightly doped n-type channel layer is then separated from the substrate by a p-type AlGaAs layer. The p-type AlGaAs serves two purposes: First, the band edge discontinuity provides the bottom part of the vertical confinement. Second, the p-type doping serves to complete the depletion of the transport layer. By using dopants to deplete the transport layer, channel thicknesses an order of magnitude larger than those obtained

50 nm		P-type 2.15×10^{17}
50 nm		N-type 1.85×10^{17}
300 nm	GaAs	N-type 1.0×10^{15}
100 nm	AlGaAs with 30% Aluminum	P-type 1.0×10^{16}
GaAs Substrate		N-type 5.0×10^{15}

Fig. 4. Epitaxial layer structure of the alternative p-type heterojunction acoustic charge transport device (p-HACT).

in current HACT devices can be achieved, sensitivity to the surface states is reduced, and processing variations can be avoided.

The remainder of the paper will discuss the two dimensional device simulation code, Semiconductor *Total Energy Balance Simulator* in two *Dimensions* (STEBBS-2D), used for modeling the performance of the HACT structures. The charge packet size and shape for the two different HACT designs at one specific acoustic power level will then be compared for three different residual surface charge concentrations. Finally, transient simulations using several values of carrier lifetime in the channel layer provide a measure of the maximum charge transfer efficiency of the proposed structure.

II. DEVICE DESIGN

The layer structure and doping concentrations in the new p-type HACT, hereafter referred to as p-HACT, is shown in Fig. 4. The design consists of a top layer of p-type GaAs doped to $2.15 \times 10^{17} \text{ cm}^{-3}$ with a thickness of 50 nm, 50 nm n-type GaAs doped to $1.85 \times 10^{17} \text{ cm}^{-3}$ follows the cap layer. The channel layer is 300 nm of n-type GaAs doped to $1.0 \times 10^{15} \text{ cm}^{-3}$. The bottom vertical confinement layer is 100 nm of p-type $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ with a doping concentration of $1.0 \times 10^{16} \text{ cm}^{-3}$. The substrate is n-type GaAs doped at $5.0 \times 10^{15} \text{ cm}^{-3}$. Alternatively a $2 \mu\text{m}$ GaAs buffer layer of the same doping concentration on a semi-insulating substrate can be used. The one dimensional conduction band diagram of the p-HACT, without the superimposed acoustic potential, for three different residual surface state concentrations is shown in Fig. 5, the surface state densities are $-5.0 \times 10^{11} \text{ cm}^{-2}$, $-1.0 \times 10^{12} \text{ cm}^{-2}$, and $-5.0 \times 10^{12} \text{ cm}^{-2}$ which differ from those used in the n-HACT simulation. Comparing Figs. 3 and 5 it is immediately apparent that the p-HACT structure is much less sensitive to fluctuations in the top surface charge; the band diagram does not change as the surface charge varies. Also note that the width of the transport layer has been increased by almost an order of magnitude, from 40 nm to 300 nm.

The maximum doping concentration in the two HACT structures is approximately the same for several reasons. First, it was found that when the doping concentrations in the first two layers was reduced the p-HACT became more sensitive to the surface state density. Similar behavior occurs for the n-HACT structure. Since the design goal is to produce a device insensitive to the surface state concentration, the doping

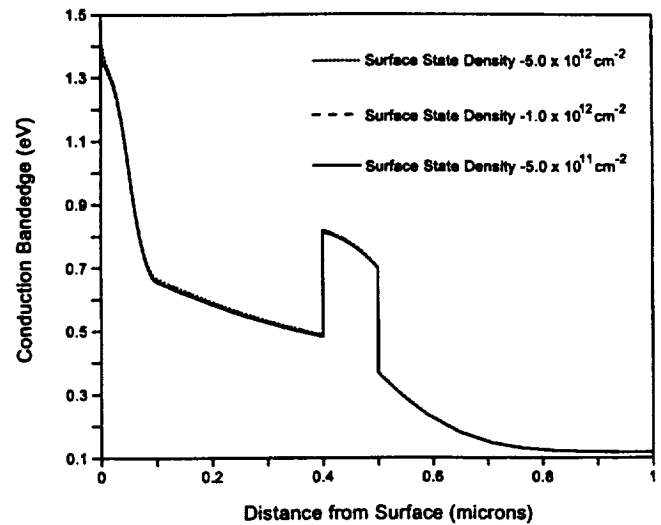


Fig. 5. One dimensional conduction band diagrams of the p-HACT device with three different residual surface state concentrations, no acoustic potential.

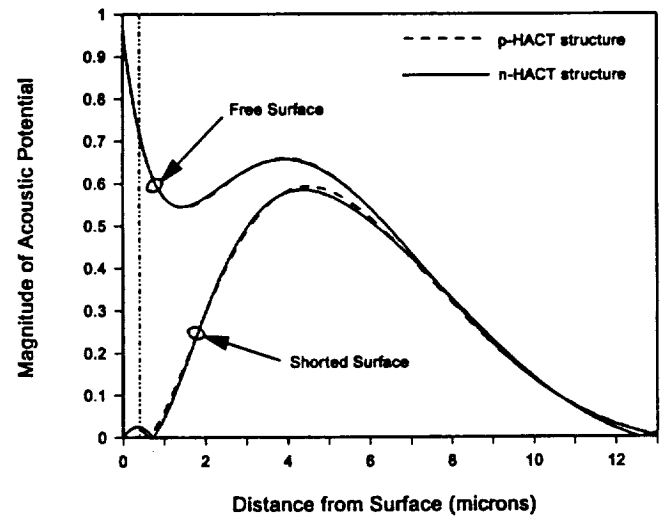


Fig. 6. Magnitude of the acoustic potential in the n-HACT and p-HACT structures, the transport channels are located between the origin and the vertical line. Two different boundary conditions are displayed; a free surface boundary condition and a shorted (metallized) boundary condition.

concentrations in the top layers must remain fairly high. However, the doping concentration in these layers can not be increased to arbitrarily high values in either structure owing to the fact that the top surface of the HACT structure can not be highly conductive [1]. A highly conductive surface, such as a metal or highly doped semiconductor, will short out the surface acoustic wave propagating in the underlying semiconductor. This will be discussed at length in the next section. In the case of the n-HACT structure the doping concentration in the AlGaAs cap layer can not be raised too high because the layer will begin modulation doping the transport layer, similar to a HEMT. As the number of carriers increases in the channel due to the modulation doping the acoustic wave will be damped out, the perturbation due to the displacement charge from the surface acoustic wave will become too small compared to the background carrier concentration. In the p-HACT device,

$$p\bar{\nu}_p = \mu_h \left[-k_b T_h \Lambda_v \nabla p - p \nabla (\epsilon_v - \chi - E_g) + \frac{3}{2} p k_b T_h \Lambda_v \frac{\nabla m_h^*}{m_h^*} \right]. \quad (5)$$

high doping creates a very strongly depleted channel but the perturbation due to the acoustic wave is not great enough to cause significant lateral confinement. For this case the heavy doping strongly pins the energy bands in the channel region.

III. MODEL DESCRIPTION

The ability to model the HACT structure depends upon the assumption that the acoustic properties of the material can be treated as a perturbation to the electrical properties under specific conditions. The SAW potential is converted to an effective doping concentration and superimposed on the background concentration of the structure. This is more realistic than superimposing the SAW potential directly on the device potential because at the surface and deep within the substrate the SAW contributes little to the overall potential.

A. Electrical Model

The fundamental basis of our model is a two dimensional hydrodynamic device simulation code, STEBS-2D [4]. For the simulations presented here, the lattice temperature effects and electronic heat transfer have been neglected. The resulting partial differential equations describing the system and solved within the context of the hydrodynamic simulator are

$$\nabla(\xi_r \nabla \psi) = \frac{q}{\xi_o} (p - n + N_d^+ - N_a^- + \Gamma_{SAW}) \quad (1)$$

$$\left(\frac{\partial n}{\partial t} \right) + \nabla \cdot (n \bar{\nu}_n) = -R + G \quad (2)$$

$$\left(\frac{\partial p}{\partial t} \right) + \nabla \cdot (p \bar{\nu}_p) = -R + G \quad (3)$$

$$n \bar{\nu}_n = \mu_n \left[-k_b T_e \Lambda_c \nabla n + n \nabla (\epsilon_c - \chi) + \frac{3}{2} n k_b T_e \Lambda_c \frac{\nabla m_e^*}{m_e^*} \right] \quad (4)$$

(see (5) shown at the top of the page.)

Equation (1) is Poisson's equation for the potential and includes the effective doping from the surface acoustic wave, Γ_{SAW} , as will be described. Equations (2) and (3) are the electron and hole current continuity equations. The electron and hole flux equations are given in (4) and (5). On the right hand side of (2) and (3), the generation/recombination terms (R and G) account for band to band, Auger, and Shockley-Read-Hall (SRH) recombination. The partial differential equations are discretized using the control volume method as described in [4], [5]. All material parameters i.e. bandgap, mobility, density of states, recombination parameters are determined using the formulas stated in PC-1D [6]. In addition, a thermionic-field emission boundary condition [7] for the currents is implemented at each material discontinuity. Interface states

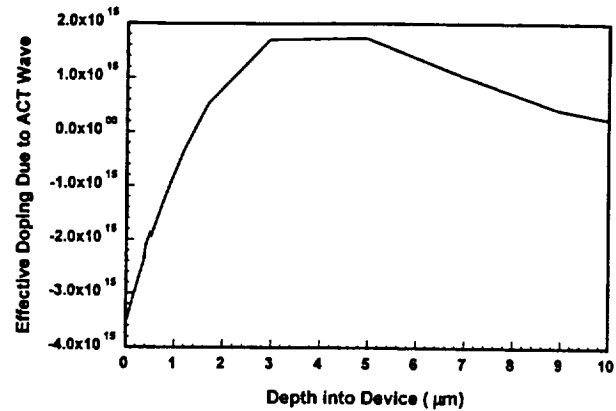


Fig. 7. Maximum value of the displacement charge created by the acoustic wave as a function of depth into the p-HACT structure. The discontinuities are caused by abrupt changes in material permittivity.

and recombination centers were assumed to be negligible at the material discontinuities. Equations (1)–(5), along with the appropriate boundary conditions for the device domain, are solved for the three fundamental variables ψ , n , and p .

In the simulations presented here four different boundary conditions are needed. The first is a grounded ohmic contact [8] at the back surface of the device. The second type of boundary condition is the symmetry boundary condition [8]. This is applied on both lateral sides of the simulation domain. This choice of boundary conditions will be explained further in the next section on modeling the acoustic wave. On the top surface two different types of boundary conditions are used. In the one dimensional simulations without the surface acoustic wave, Figs. 3 and 5, a free surface boundary condition is employed for the electrostatic potential. Its mathematical form is

$$\epsilon_{semi} \frac{\partial \psi}{\partial X_i} = Q_{int} \quad (6)$$

The parameter ϵ_{semi} is the semiconductor permittivity and Q_{int} is the surface charge density and varies in the range -5×10^{11} to $-5 \times 10^{12} \text{ cm}^{-2}$. This range of surface charge density corresponds to approximately 1% of the dangling bonds being charged. Even though the SAW is expected to alter the surface state occupancy (Q_{int}), this has been neglected in the model due to the lack of information concerning the charged state profile. The boundary conditions for the carrier concentrations depend upon satisfying the continuity equations at the surface with respect to the surface recombination rate as,

$$J_n = -q R_{surf} = -q \frac{S_n S_p (np - n_i^2)}{S_n (n + n_1) + S_p (p + p_1)} \cdot \delta(x) \quad (7)$$

$$J_p = q R_{surf} = q \frac{S_n S_p (np - n_i^2)}{S_n (n + n_1) + S_p (p + p_1)} \cdot \delta(x)$$

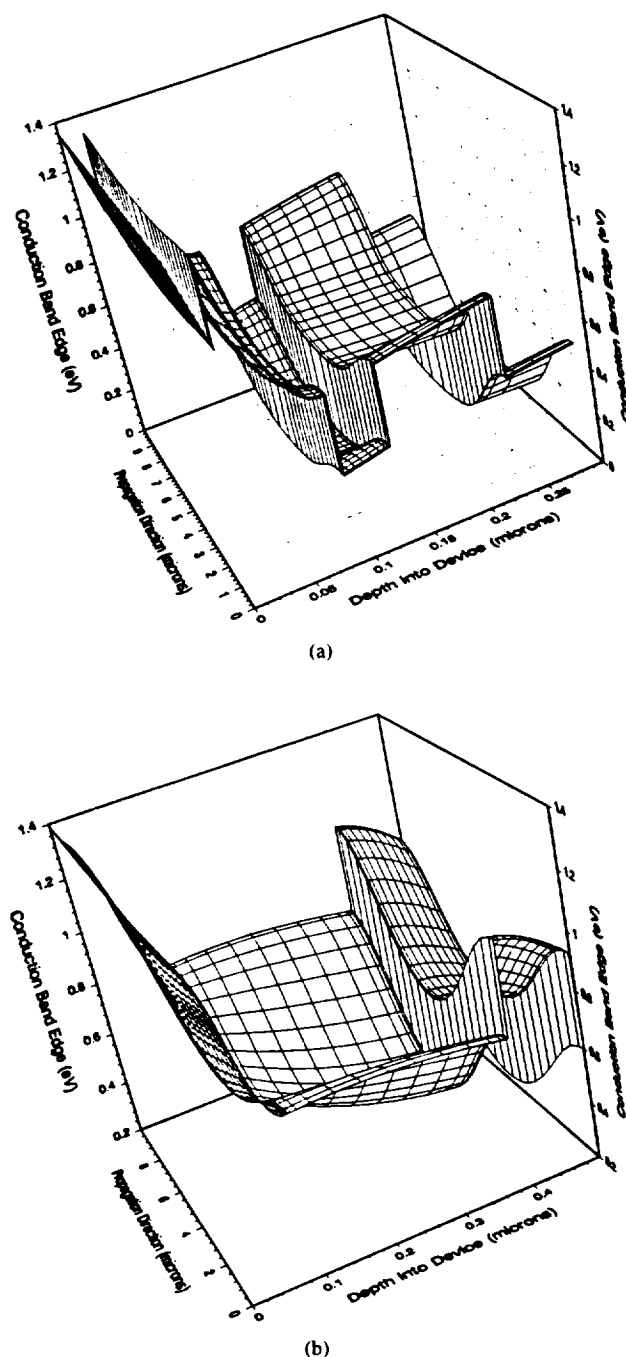


Fig. 8. Two dimensional conduction band diagrams for the (a) n-HACT structure, and (b) p-HACT structure with an applied surface acoustic wave. Only the epitaxial layer are displayed in each case.

In the one dimensional simulations the value of the surface recombination rates (R_{surf}) used in (7) has been set to zero. Setting R_{surf} to zero is justified because the surface is not driven out of equilibrium. Therefore $np = n_i^2$ and no matter what the actual surface recombination velocities are the surface recombination rate evaluates to zero [8]. These one dimensional simulations then provided the value of the fermi level with respect to the intrinsic level which was used in the two dimensional simulations.

In the two dimensional simulations with the SAW the free surface boundary conditions given by (6) and (7) in the

one dimensional case could be used with the modification that an extra term accounting for the acoustic displacement charge is added to (6). This is only done in the case of the n-HACT structure at the lowest residual surface state concentration. At the other surface state concentrations in the n-HACT structure and at all surface state concentrations in the p-HACT structure the free surface boundary condition is replaced by a pinned fermi level boundary condition. The value of the fermi level is fixed to the value calculated from the one dimensional simulation with the fixed surface charge. The use of either the free surface with charge boundary condition or the pinned fermi level boundary condition in these cases results in the same potential profile. The similar potential profile occurs because the SAW displacement charge is a very minor perturbation compared to the fixed surface states, only changing the fermi level by $\sim 1.0 \times 10^{-5}$ volts. Even in the n-HACT structure at the lowest surface state concentration and a modified free surface condition of (6) with the acoustic displacement, the potential profile within the channel is unchanged from that obtained with a fixed fermi level at the surface. However, the pinned fermi level boundary condition tends to speed the convergence of the solution.

The simulation domain is nonuniformly discretized using rectangular control volumes. The node placement is performed adaptively during the equilibrium portion of the simulation. The adaptive gridding depends upon the change in potential, changes in material types or boundary conditions, and fluctuations in the doping concentrations. Generally the adaptive gridding results in a mesh consisting of approximately 107×41 points. The system of equations is linearized using Newton's method. The resulting system is solved using an iterative approach known as the conjugate gradient squared and stabilized method [9]. The code for this accelerator was added to the NSPCG library for the solution of nonsymmetric systems of equations [10].

B. Modeling the Surface Acoustic Wave

As previously stated the acoustic wave potential provides lateral confinement and transports the charge. To utilize the maximum coupling coefficient between the mechanical and electrical components of the SAW power a wave propagating in the $\langle 110 \rangle$ direction on a $[100]$ crystal structure is simulated. The SAW potential is calculated for the layer structures by the technique described in reference [11]. The resulting normalized potential as a function of distance from the surface for the p-HACT structure supporting a SAW with a power intensity of $1 \text{ mWatt}/\lambda$ is depicted in Fig. 6. The magnitude of the SAW potential for two different surface conditions, a free surface and a metallized surface, are displayed. Notice that the SAW potential is completely shorted out at the surface for the metallized device while the SAW potential maximum occurs at the surface in the free surface structure. Fig. 6 can be used to explain why a highly conductive surface can not be used in the HACT structure, as mentioned above. The location of both the n-HACT and p-HACT channels lie between the origin and the vertical line in Fig. 6. The channels are located very close to the surface to take advantage of the high magnitude of the

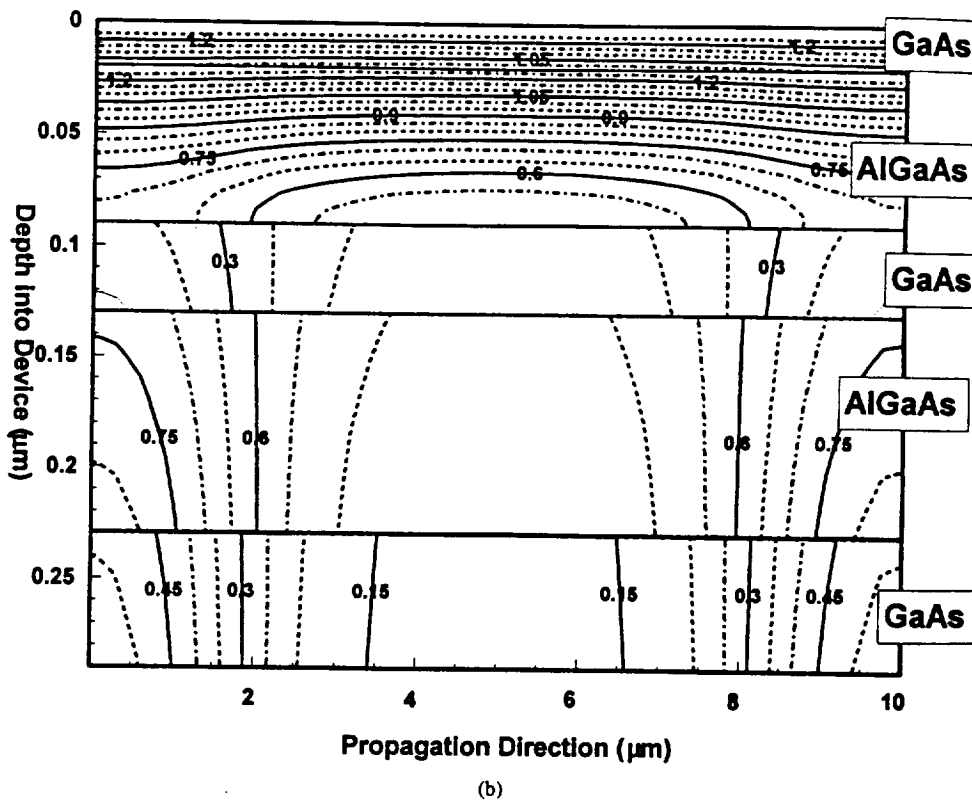
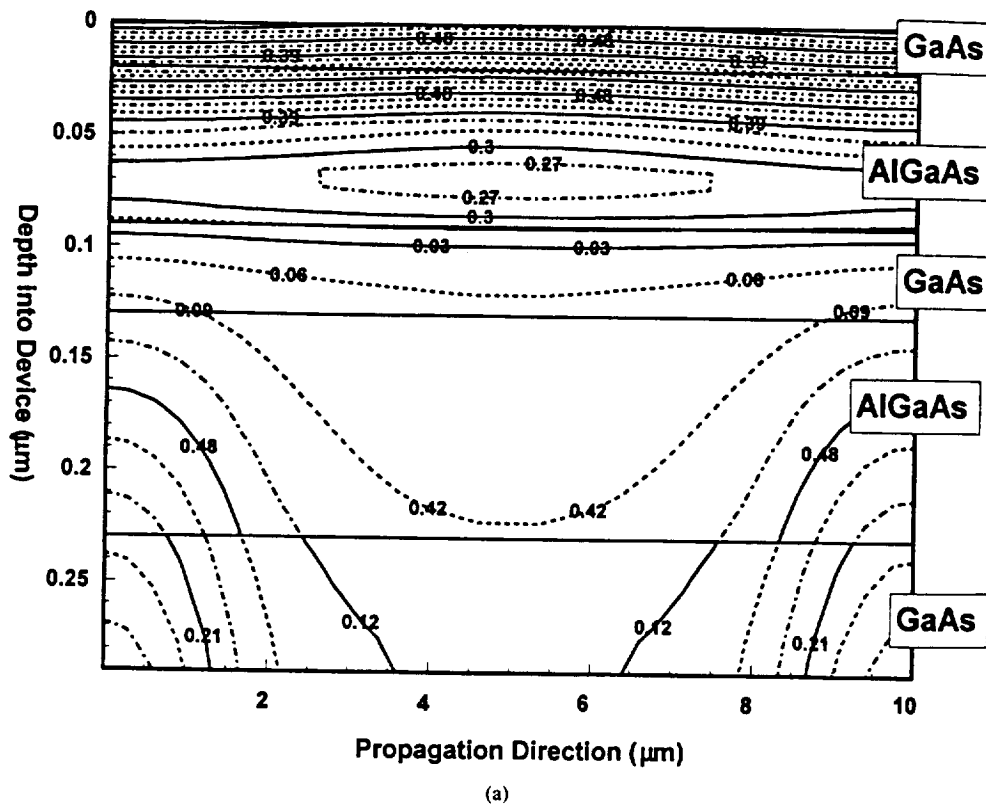


Fig. 9. Equi-energy contour plots of the n-HACT structure with an applied surface acoustic wave and three different residual surface state concentrations: (a) $-1.0 \times 10^{12} \text{ cm}^{-2}$, (b) $-1.5 \times 10^{12} \text{ cm}^{-2}$.

SAW potential present in the free surface case. As the front surface doping concentration increases, the surface becomes conductive and the maximum magnitude of the SAW potential

decreases in the immediate vicinity of the surface. Therefore, in the highly conductive surface device the SAW potential is insufficient near the surface to effectively transport charge.

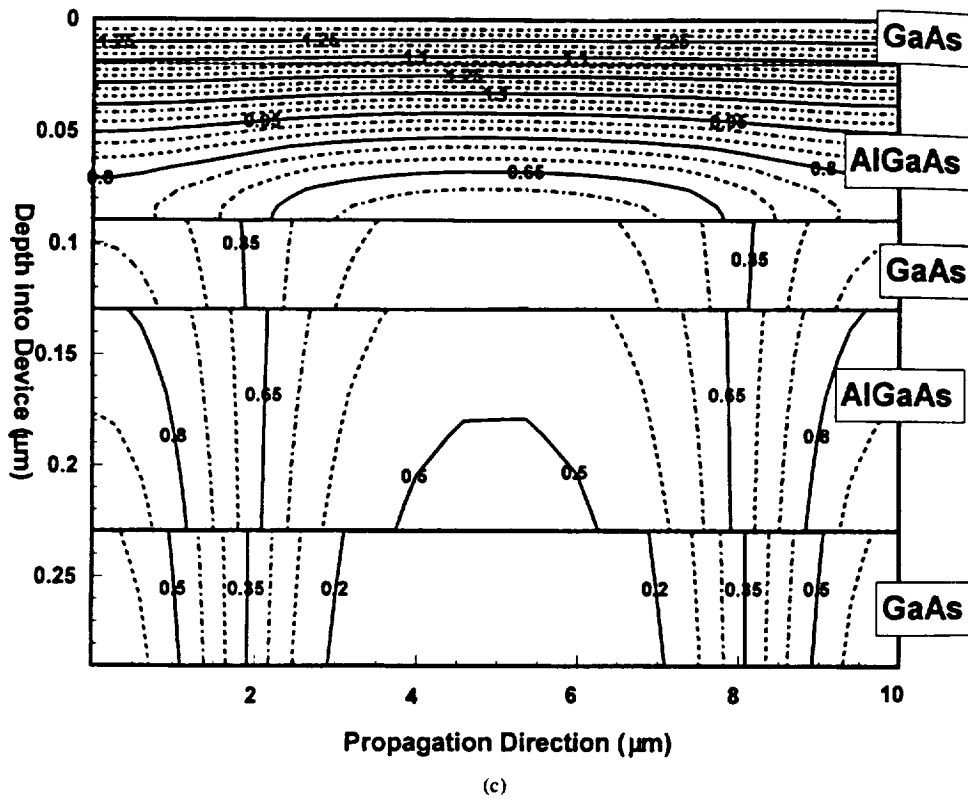


Fig. 9. (c) $-5.0 \times 10^{12} \text{ cm}^{-2}$.

As stated in Section III.B. the acoustic wave is modeled as a perturbation to the impurity doping density in Poisson's equation. The magnitude of the displacement charge resulting from the acoustic wave is obtained by the Laplacian of the wave potential and modulated by a sinusoidally varying function [12]

$$\Gamma_{saw}(x, y, t) = P_s \frac{\epsilon}{q} \left[k^2 \phi_{saw}(x) - \frac{\partial^2 \phi_{saw}(x)}{\partial x^2} \right] \cos(ky - \omega t) \quad (8)$$

The factor P_s in (8) is a power scaling factor, indicating the SAW potential at the surface under free surface conditions, while k is the wave number of the acoustic wave. Fig. 7 depicts the maximum displacement charge, $\Gamma_{SAW}(x, y, t)$, in the p-HACT structure for an acoustic wavelength of $10 \mu\text{m}$, a P_s of 4 [12], and a free surface boundary condition. Contrary to the smooth behavior of the displacement charge in the case of bulk ACT devices the displacement charge in the HACT case shows discontinuities due to the abrupt changes in the permittivity of the constituent material.

In the calculations presented below only one wavelength of the SAW will be simulated. The beginning and ending points are chosen at the potential minimums, thus creating a potential well to confine the electrons in the middle of the simulation domain. Provided the injected electrons do not greatly effect the shape of the SAW the choice of a symmetry boundary condition in the lateral directions can be justified.

C. Comparison with Previous Models

Several numerical models have been reported to simulate charge injection in bulk ACT devices [12]–[15]. One of the

major aims of this work is to analyze the HACT structures so only a limited comparison of our model to previous ACT models is appropriate. The earlier bulk ACT models neglect the hole contribution to the solution. This implies that no carrier recombination takes place, and therefore the charge transfer efficiency would always be calculated to be unity. Our model enables a more realistic calculation of the charge transfer efficiency by including the hole transport equation to simulate recombination. Previous ACT models were concerned with simulating the injection process between the input ohmic contact on the surface and the acoustic transport channel in the bulk of the device. In the model presented here the desired amount of charge is added at a specific location very close to the surface in a very small time step (1 psec). Then two small time steps are sequentially performed allowing the charge to distribute itself within the SAW's well. The time steps used in the simulation are 0.1 and 0.5 nsec. After 0.6 nsec elapses the injected charge is then in steady state with the acoustic potential. Next the time steps are increased to calculate the charge transfer efficiency.

Simulation of an n-HACT structure has been reported by researchers at United Technologies Research Center (UTRC) [2], [16]. As in the case of the bulk ACT models the hole equation was removed from the UTRC solution domain. The charge injection in the UTRC model is similar to the method of injection adopted here. A major difference between the two models is the UTRC model uses top and bottom ground planes as boundary conditions for the top and bottom of the device. It is in these planes that image charge created by the injected electrons will reside. Using the model presented here, which

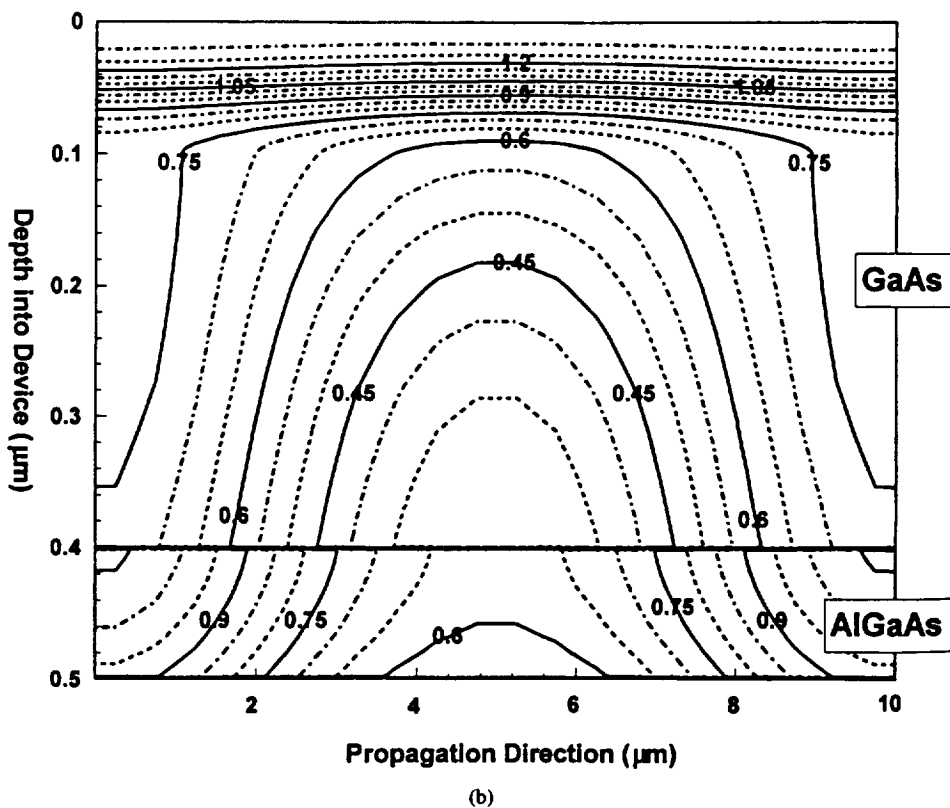
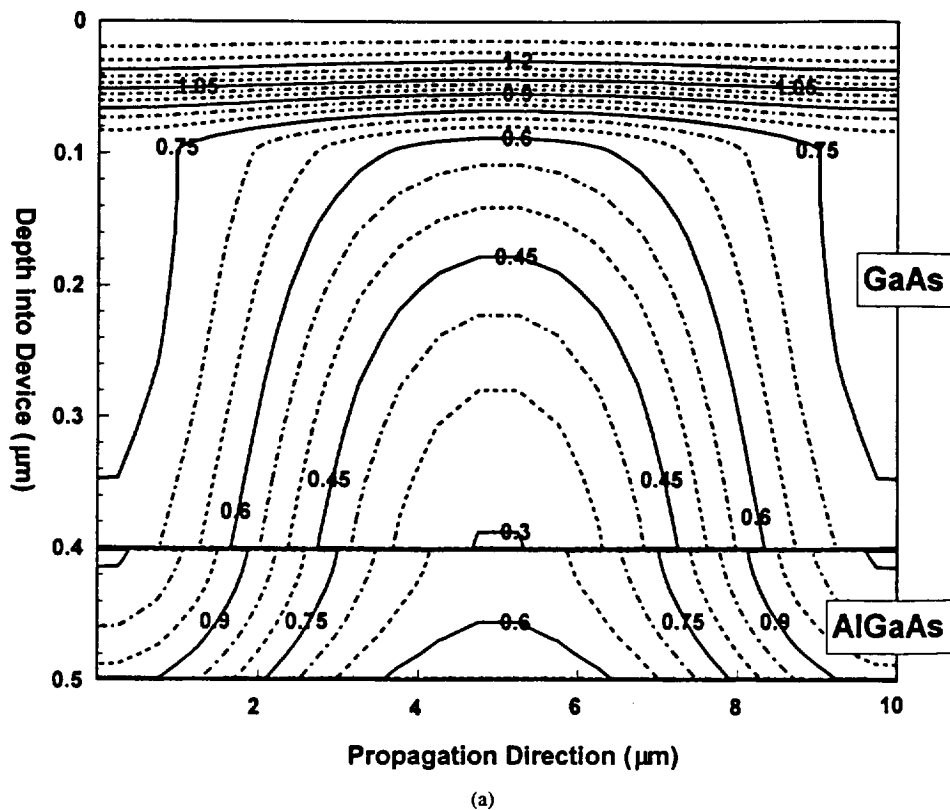


Fig. 10. Equi-energy contour plots of the p-HACT structure with an applied surface acoustic wave and three different residual surface state concentrations: (a) $-5.0 \times 10^{11} \text{ cm}^{-2}$, (b) $-1.0 \times 10^{12} \text{ cm}^{-2}$.

does not use ground planes, it will be shown that the image charge actually resides in the substrate just below the bottom

heterojunction. Another major difference is in the calculation of the displacement charge. The UTRC model used a simple

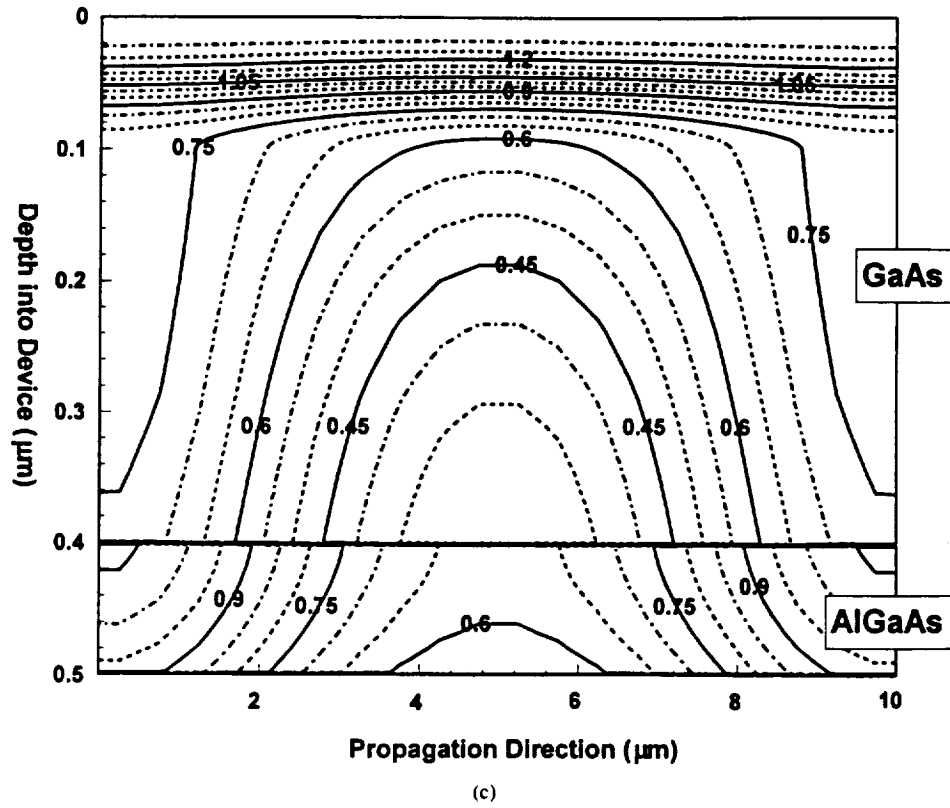


Fig. 10. (c) $-5.0 \times 10^{12} \text{ cm}^{-2}$.

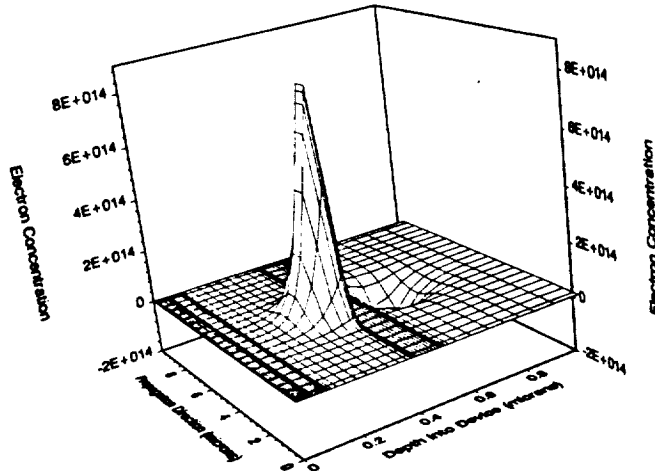


Fig. 11. Electron packet shape for the p-HACT structure with an applied surface acoustic wave and 10^6 electrons/cm injected into the transport channel. The loss of electrons from 0.6 to 0.8 μm deep is the image charge created by the injected electrons.

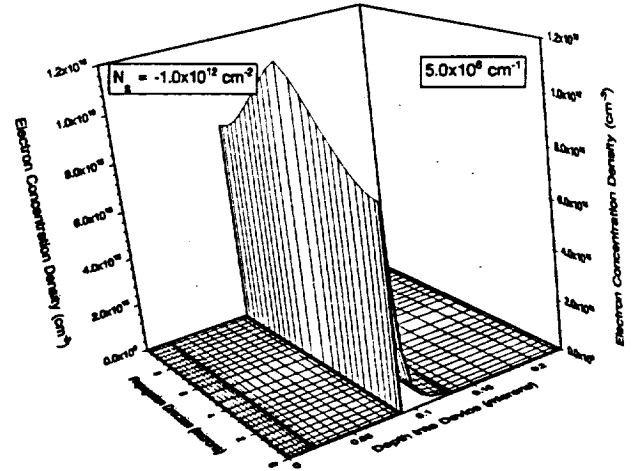


Fig. 12. Electron packet shape for the n-HACT structure with an applied surface acoustic wave, a residual surface state concentration of $-1.0 \times 10^{12} \text{ cm}^{-2}$, and an injected carrier concentration of 5×10^6 electrons/cm.

expression for the displacement charge in the HACT channel,

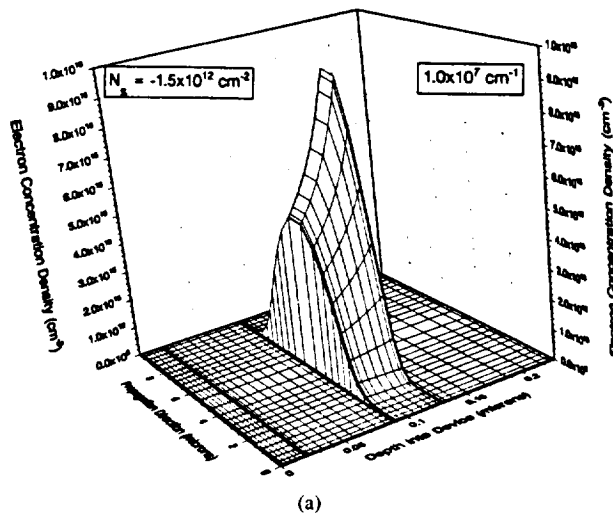
$$\Gamma_{saw}(x, y, t) = -\left(\frac{\epsilon}{4\pi q}\right) \nabla^2(\phi_o \cos(ky - \omega t)) \quad (9)$$

where the constant ϕ_o was set to 1. In the model presented here, which uses (8), the value of the power parameter, P_s , would need to be set to an unusually high number as compared to that given in [12] to obtain the same potential profile as shown in [16]. The UTRC model and the model presented

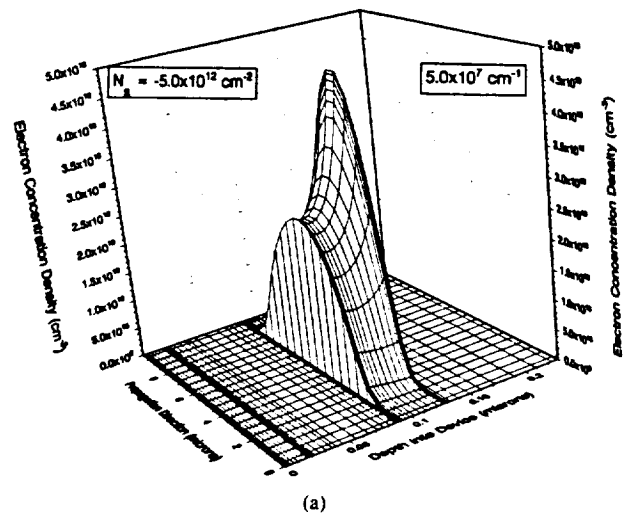
in this work both ignore the synchronous field created by the constant velocity of the SAW in the propagation direction [1].

IV. APPLICATION TO n-HACT AND p-HACT DEVICES

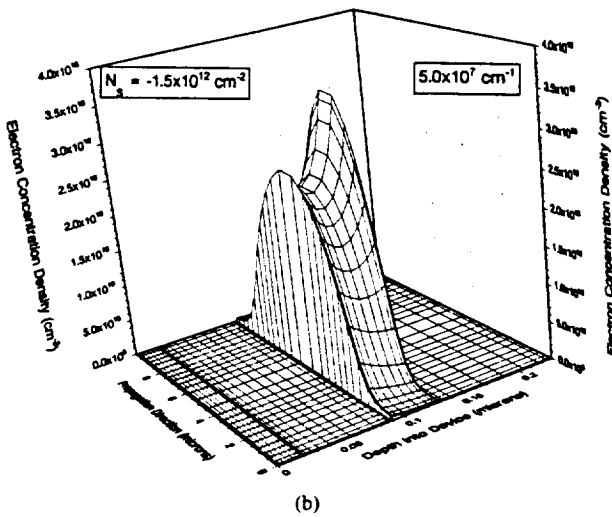
The one dimensional simulations, Figs. 3 and 5, depict the conduction band without any acoustic potential. These figures serve to highlight the effect of the surface state density for both the n-HACT and p-HACT structures. In this section these structures are simulated in two dimensions



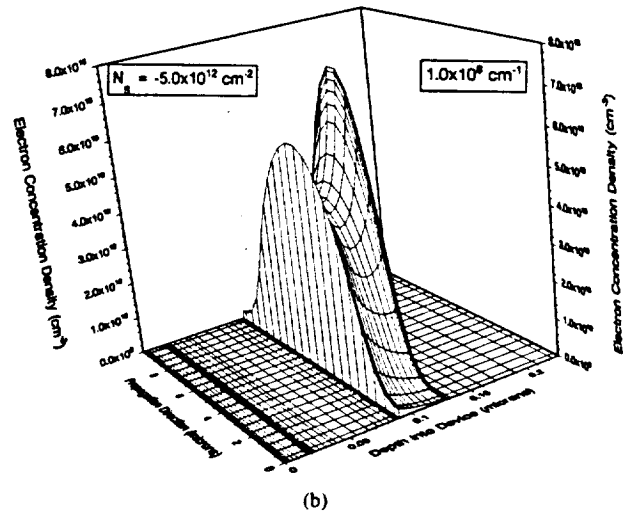
(a)



(a)



(b)



(b)

Fig. 13. Electron packet shape for the n-HACT structure with an applied surface acoustic wave, a residual surface state concentration of $-1.5 \times 10^{12} \text{ cm}^{-2}$, and two injected carrier concentrations: (a) 1.0×10^7 electrons/cm and (b) 5.0×10^7 electrons/cm. Notice the spreading in the case of $5. \times 10^7$ electrons/cm.

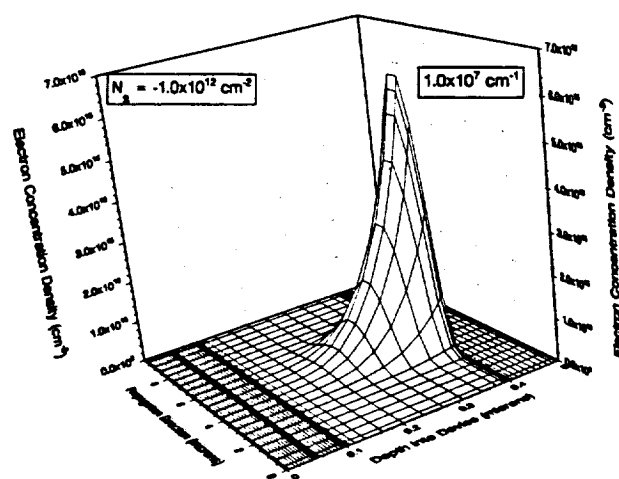
Fig. 14. Electron packet shape for the n-HACT structure with an applied surface acoustic wave, a residual surface state concentration of $-5.0 \times 10^{12} \text{ cm}^{-2}$, and two injected carrier concentrations: (a) 5.0×10^7 electrons/cm and (b) 1.0×10^8 electrons/cm.

with a superimposed SAW potential and injected charge to determine their charge capacity and charge transfer efficiency. The acoustic wave simulated has a wavelength of $10 \mu\text{m}$, a frequency of 287.75 MHz, which leads to an acoustic velocity of 2877.5 m/sec.

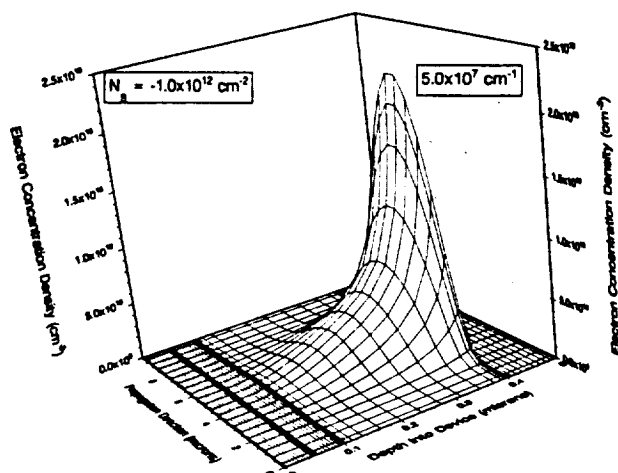
Fig. 8(a) and (b) show energy surface plots of the steady state conduction band for the n-HACT and p-HACT structures with a surface state density of $-1.5 \times 10^{12} \text{ cm}^{-2}$ and $-1.0 \times 10^{12} \text{ cm}^{-2}$ respectively. The power scaling factor, P_s , for both structures was 4. These figures are useful for illustrating the potential well created by the SAW potential. Hereafter, the conduction band will be illustrated by using contour plots of the energy surface. Comparison of Fig. 8(a) and (b) shows that the well created in the p-HACT structure (Fig. 8(b)) is approximately an order of magnitude larger than the n-HACT structure. In the n-HACT structure the top and bottom vertical confinement is provided by the two heterojunction barriers. As previously described, the top vertical confinement in the p-HACT structure is provided by the pn junction, with no

abrupt change in the conduction band, and the bottom vertical confinement is formed by the heterojunction discontinuity.

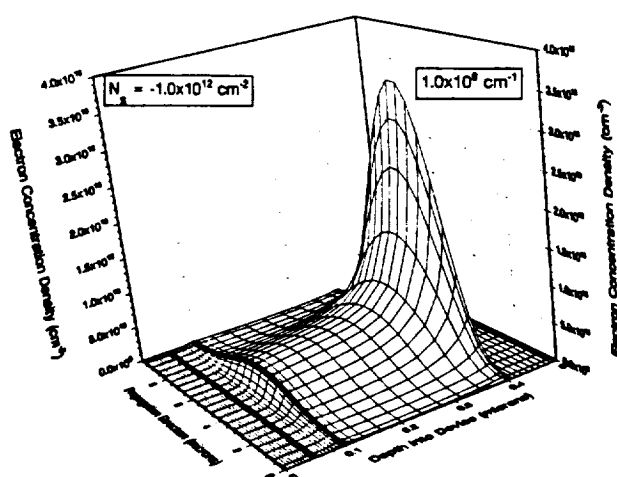
Fig. 9(a)–(c) show the equi-energy contour levels for the n-HACT structure with a P_s factor of 4 and surface state concentrations of -1×10^{12} , -1.5×10^{12} , and $-5.0 \times 10^{12} \text{ cm}^{-2}$. The values of the lateral barrier created by the SAW and the value of the pinned fermi level at the front surface are given in Table I. Note that there exists only a very little lateral potential barrier at a surface concentration of $-1.0 \times 10^{-12} \text{ cm}^{-2}$. The value of the pinned fermi level in Table I for the n-HACT structure demonstrates again this structure's sensitivity to the surface state density. At a residual concentration of $-1.0 \times 10^{-12} \text{ cm}^{-2}$ the level is 0.0895 eV, both the cap layer and channel layer are modulation doped by the AlGaAs. However, at residual concentration of $-1.5 \times 10^{-12} \text{ cm}^{-2}$ the level has dropped to -0.684 eV and the cap layer is now depleted. The steady state conduction band contours for the p-HACT structure with surface concentrations of -5×10^{11} , -1.0×10^{12} , and $-5.0 \times 10^{12} \text{ cm}^{-2}$ are displayed in Fig. 10(a)–(c). As in the case of the n-HACT structures the P_s value was



(a)



(b)



(c)

Fig. 15. Electron packet shape for the p-HACT structure with an applied surface acoustic wave, a residual surface state concentration of $-1.0 \times 10^{12} \text{ cm}^{-2}$, and three injected carrier concentrations: (a) 1.0×10^7 electrons/cm, (b) 5.0×10^7 electrons/cm, and (c) 1.0×10^8 electrons/cm. Notice the spreading in the case of 1.0×10^8 electrons/cm is much less than in the n-HACT structure Fig. 14.

TABLE I

n-HACT structure			
Surface States (cm^{-2})	-1.0×10^{12}	-1.5×10^{12}	-5.0×10^{12}
Pinned Fermi Level (eV)	0.0895	-0.6837	-0.8140
Lateral Barrier at Top of Channel (KT)	0.37	7.51	10.09
Lateral Barrier at Bottom of Channel (KT)	1.64	9.83	13.30
p-HACT structure			
Surface States (cm^{-2})	-5.0×10^{11}	-1.0×10^{12}	-5.0×10^{12}
Pinned Fermi Level (eV)	-0.6953	-0.7254	-0.8243
Lateral Barrier at Top of Channel (KT)	8.02	7.92	7.80
Lateral Barrier at Bottom of Channel (KT)	16.68	16.66	16.64

set to 4 and values of lateral barrier and pinned fermi level are listed in Table I. As seen from Fig. 10 and the values of the lateral barrier in Table I the surface concentration does not significantly change the shape of the potential well in the p-HACT structure.

To determine the maximum charge capacity of the potential well excess electrons are injected into the channel at the conduction band minimum. The structures are simulated until the electrons reach steady state with the acoustic potential. The number of injected electrons is increased until it appears that the electrons are not being confined within a single wavelength. Fig. 11 is an example of the electron charge packet shape when 10^6 electrons/cm are injected into the middle of the simulation domain for the p-HACT structure. To determine the electron packet shape the electron density simulated at steady state was subtracted from the electron density simulated 0.6 nanoseconds after the charge was injected. The 0.6 nanoseconds is required for the electrons which are only injected at the middle to attain a steady state condition with the acoustic potential. Notice that the electrons are confined to the center of the simulation domain, which is the lowest part of the conduction band in Fig. 10. The width of the packet is approximately $2 \mu\text{m}$ wide in the propagation direction. As will be seen in the figures to be presented the width of the packet increases as the amount of injected charge increases. Also note in Fig. 11 that an image charge, i.e. removal of electrons, has been created in the substrate just below the confining heterojunction. This image charge will not be plotted in later figures, but the reader should be aware that it exists.

Fig. 12 shows the electron packet shape for 5×10^6 electrons/cm injected into the n-HACT structure with the lowest residual surface state concentration. The electrons are not confined to the middle of the simulation domain, but spread over the channel along the propagation direction. This indicates that this structure has exceeded its maximum capacity for SAW transport; there is no confinement in the propagation direction. Fig. 13 depicts the electron packet shape for the n-HACT structure with a residual surface state concentration of $-1.5 \times 10^{12} \text{ cm}^{-2}$ for two injected electron concentrations. Fig. 13 a is for a packet of 10^7 electrons/cm. This is the maximum amount predicted

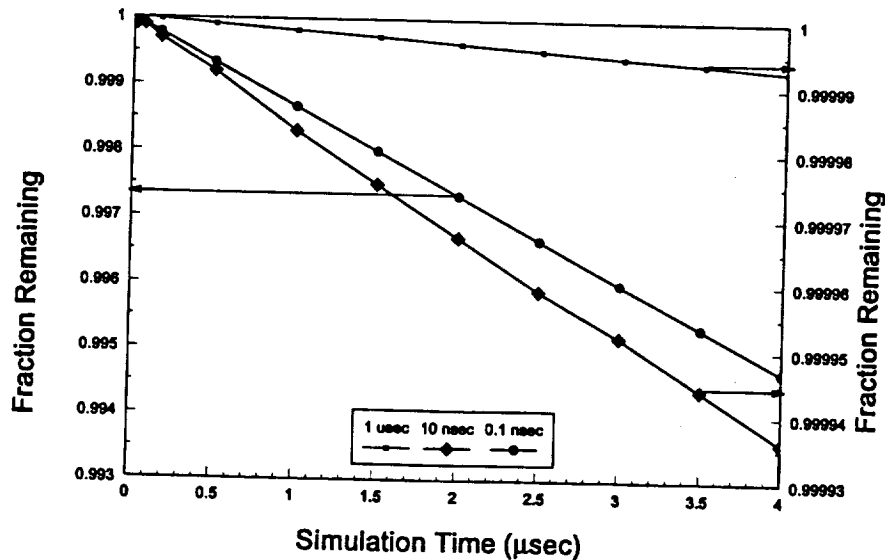


Fig. 16. Fraction of packet electrons remaining as a function of elapsed time in the p-HACT structure for three different values of SRH lifetime: 1 μsec, 10 nsec, and 0.1 nsec.

by the UTRC model for this structure. When the injected electron concentration is increased to 5×10^7 electrons/cm, Fig. 13(b), there is spreading of the packet over the entire simulation domain, indicating that the maximum capacity has been exceeded. Note that unlike the UTRC model the lack of confinement is due to the SAW potential not the heterojunction barrier. Fig. 14 displays the electron packet shape for the n-HACT structure with the highest residual surface state concentration. In this case the electron packet does not spread until the injected concentration reaches 10^8 electrons/cm. Comparison of Figs. 13 and 14 shows that excess surface charge can lead to a small increase in the maximum charge capacity of the wells.

The electron packet shape obtained using the p-HACT structure with a residual surface state concentration of $-1.0 \times 10^{12} \text{ cm}^{-2}$ and three different injected electron concentrations is shown in Fig. 15. For an injected electron concentration of 10^7 electrons/cm, Fig. 15(a), the packet is confined within $5 \mu\text{m}$ of the simulation domain. Fig. 15(b) shows the packet confinement when 5×10^7 electrons/cm are injected, even at 10^8 electrons/cm, Fig. 15(c), the packet is still fairly well confined to the simulation domain. Due to the similarities in the characteristics of the p-HACT structure at other residual surface state densities charge packet shapes at other surface state concentrations will not be presented. Simulations at surface state densities of 1.0×10^{11} and $1.0 \times 10^{13} \text{ cm}^{-2}$ also show similar insensitivity. Comparison of Figs. 13 and 15 shows that due to the larger channel thickness in the p-HACT structure the peak electron concentration density is approximately half of that obtained in the n-HACT structure for the same injected carrier concentration. These figures also show that for approximately the same surface concentration the p-HACT has a charge capacity which is nearly an order of magnitude higher than the n-HACT structure. The maximum charge capacity for the p-HACT structure is approximately the same as the n-HACT structure with a high residual surface

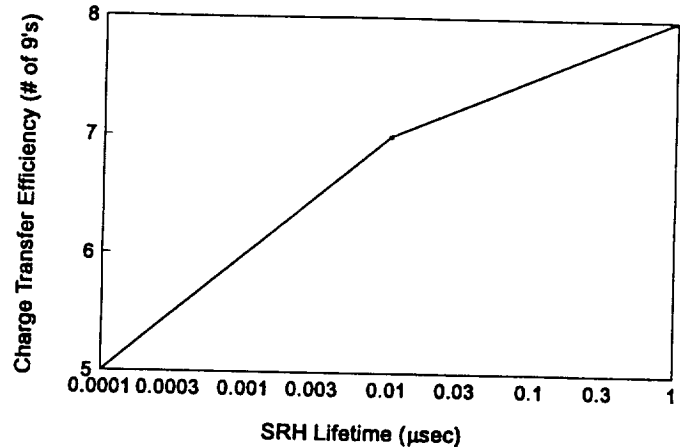


Fig. 17. Number of 9's in the charge transfer efficiency as a function of the value of the SRH lifetime in the channel region of the p-HACT device structure.

concentration, as can be seen from a comparison of Figs. 14 and 15.

To determine the maximum charge transfer efficiency of the p-HACT structure electrons are injected into the channel layer in the same manner as for the packet shape calculations. The structure is simulated for ~ 1000 wavelengths, $4 \mu\text{sec}$, at different time intervals. The amount of electrons remaining in the packet was recorded to determine the transfer efficiency. The fraction of electrons remaining in the packet as a function of time is shown in Fig. 16 for three different values of SRH lifetime in the channel layer. At high SRH lifetimes the fraction of remaining electrons is very close to unity, indicating excellent charge transfer efficiency. The calculated charge transfer efficiency, given by the number of 9's in efficiency, as a function of SRH lifetime is given in Fig. 17. For a SRH lifetime of 0.1 nsec the structure has 5 9's transfer efficiency and increases to 8 9's for 1 μsec lifetime. However, this analysis does not include any interface recombination at

the heterojunction nor does it account for slow traps in the channel layer. Both of these effects will act to lower the charge transfer efficiency.

V. CONCLUSIONS

An alternative HACT structure which uses a $pn^{-1}np$ doping profile to deplete the transport layer and isolate the channel from the residual surface states has been analyzed. The $pn^{-1}np$ doping also leads to an increase in the channel thickness. The p-HACT structure is insensitive to the residual surface state density over two orders of magnitude. Charge capacity simulations show that the p-HACT structure is capable of nearly an order of magnitude improvement in electron capacity over the n-HACT structure at nearly the same residual surface concentration. Simulations also show that the charge transfer is 5 to 8 9's efficient if the recombination is limited by SRH lifetimes in the range of 0.1 nsec to 1 μ sec. The p-HACT structure described in this paper has not been fully optimized, further calculations reveal that the doping in the first n-type layer may need to be reduced.

REFERENCES

- [1] R. L. Miller, C. E. Nothnick, and D. S. Bailey, "Acoustic Charge Transport: Device Technology and Applications," Artech House, Inc. Norwood, MA, 1992.
- [2] G. A. Peterson, B. J. McCartin, W. J. Tanski, and R. E. LaBarre, "Charge confinement in heterojunction acoustic charge transport devices," *Appl. Phys. Lett.*, vol. 55, pp. 1330-1332, 1989.
- [3] T. W. Grudkowski, R. N. Sacks, "HACT structure with reduced surface state effects," U.S. Patent #5264717, 1993.
- [4] A. W. Smith, "Light confinement and hydrodynamic modeling of semiconductor structures by volumetric techniques," Ph.D. Dissertation, Georgia Institute of Technology, 1992.
- [5] S. V. Pantankar, "Numerical Heat Transfer and Fluid Flow," Hemisphere Publishing Corp., New York, NY, 1980.
- [6] P. A. Basore, "PC-1D installation manual and user's guide version 3.0," Sandia National Laboratories, Albuquerque, NM, 1991.
- [7] K. Horio and H. Yanai, "Numerical modeling of heterojunctions using thermionic emission current at the heterojunction interface," in *Proc. of the 6th Inter. NASECODE Conf.*, pp. 390-395, Dublin, Ireland, July 11-14, 1989.
- [8] S. Selberherr, "Analysis and Simulation of Semiconductor Devices," Springer-Verlag, Wien, New York, 1984.
- [9] P. Chin and P. A. Forsyth, "A comparison of GMRES and CGSTAB accelerations for incompressible Navier-Stokes problems," *J. Comp. and Appl. Math.*, vol. 46, pp. 415-426, 1993.
- [10] T. C. Oppé, W. D. Joubert, and D. R. Kincaid, "An overview of NSPCG: a nonsymmetric preconditioned conjugate gradient package," *Computer Physics Communications*, vol. 53, pp. 283-293, 1989.
- [11] Y. Kim and W. D. Hunt, "Acoustic fields and velocities for surface-acoustic-wave propagation in multilayered structures: An extension of the Laguerre polynomial approach," *J. Appl. Phys.*, vol. 68, pp. 4993-4997, 1990.
- [12] E. G. Bogus, "Electrical charge injection in an acoustic charge transport device," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 1987.
- [13] E. G. Bogus, M. J. Hoskins, and B. J. Hunsinger, "Numerical model for electrical charge injection in the acoustic charge-transport device," *IEEE Trans. Electron Devices*, vol. 38, pp. 822-830, 1991.
- [14] S. M. Knapp, J. J. Liou, and D. C. Malocha, "Modeling the charge injection process in acoustic charge transport devices," *IEEE Ultrasonics Symp. Proc.*, pp. 223-227, 1989.
- [15] S. M. Knapp, D. C. Malocha and J. J. Liou, "A simplified and efficient numerical model for charge injection in acoustic charge transport devices," *IEEE Trans. Electron Devices*, vol. 39, pp. 1811-1820, 1992.
- [16] W. J. Tanski, D. E. Cullen, S. W. Merritt, R. N. Sacks, and G. A. Peterson, "Heterojunction acoustic charge transport devices for electronic warfare applications," Final Contractor Report to Air Force Aeronautical Laboratories #AFWAL-TR-88-1073, 1988.

Arlynn W. Smith received the B.S. degree in ceramic engineering from Alfred University, Hornell, NY, in 1984, the M.S. degree from the Georgia Institute of Technology, Atlanta, GA, in 1987, and the Ph.D. degree in electrical engineering from Georgia Tech in 1992.

From 1992 to 1993, he held a post-doctoral fellowship at the Georgia Tech Research Institute, where he developed a three-dimensional simulation code for modelling the chemical vapor infiltration of ceramic matrix composites. Currently, he is performing a post-doctoral research program with the microelectronics research center at Georgia Tech, developing a hydrodynamic simulation code of heterojunction acoustic charge transport devices, charge transport devices, and avalanche photodiodes.

J. Stevenson Kenney was born in St. Louis, MO, in 1962. He received the B.S. in electrical engineering in 1985, and the M.S. in electrical engineering in 1990, both from the Georgia Institute of Technology, Atlanta, GA. He is currently enrolled at Georgia Tech, pursuing the Ph.D. in electrical engineering.

He has over seven years of industry experience in microwave circuit and subsystem design, and has been employed by Electromagnetic Sciences, Inc., Scientific Atlanta and SPC Electronics America, Inc., all of Norcross, GA. He is currently employed by Pacific Monolithics, Inc., Sunnyvale, CA.

He has served on the steering for the 1993 International Microwave Symposium and received the Best Student Paper at the 1993 IEEE MTT-S International Microwave Symposium.



William D. Hunt was born in Jackson, MS, on December 21, 1954. He received the B.S. in 1976 from the University of Alabama, Tuscaloosa, AL, the M.S. from the Massachusetts Institute of Technology, Cambridge, MA, in 1980 and the Ph.D. from the University of Illinois at Champaign-Urbana in 1987, all in electrical engineering.

He worked for the Harris Corporation from 1976 to 1978, and for Bolt Beranek and Newman from 1980 to 1984. He joined the faculty of Georgia Tech in 1987, where he is currently an Associate

Professor. He was a Rhodes Scholar Finalist in 1975, received a DuPont Young Faculty Award in 1988 and was named an NSF Presidential Young Investigator in 1989. In addition, he has been selected for Who's Who in the South and Southwest and Who's Who in American Education and was named as one of the 1994 Distinguished Engineering Fellows by the University of Alabama College of Engineering.

His research interests include SAW and ACT devices, as well as transducers for biomedical ultrasound.



Kevin F. Brennan (S'84-M'84-SM'90) received the B.S. degree in physics from the Massachusetts Institute of Technology, Cambridge, MA, in 1978, and the M.S. degrees in physics and the Ph.D. degree in electrical engineering in 1980 and 1984, respectively, from the University of Illinois at Urbana-Champaign.

He is currently a Professor of Electrical and Computer Engineering at the Georgia Institute of Technology, Atlanta, GA. His current research interests include the theory and modeling of avalanche

devices and confined state photomultipliers, the physics of impact ionization, nonlinear transport effects in semiconductors and heterostructures, plasmons in semiconductors, and the theory and modeling of acoustic charge transport devices.

Dr. Brennan was awarded a Presidential Young Investigator Award through the National Science Foundation in 1988.

Rudy Benz was born in Detroit, MI, on February 22, 1962. He received the B.S. degree with honors in engineering science from the Pennsylvania State University, University Park, PA, in 1984, and the M.S. and Ph.D. degrees in physics from the Georgia Institute of Technology, Atlanta, GA, in 1988 and 1992, respectively.

He is currently a research scientist at the Georgia Tech Research Institute. His research interests include the growth and characterization of II-VI and III-V semiconductors by molecular beam epitaxy, the surface kinetics of epitaxial growth, and the application of GaAs charge transfer devices to solid state imager applications.

Christopher J. Summers (M'82) was born in Oxford, England, on August 3, 1940. He received the B.S. and Ph.D. degrees in physics from Reading University, Reading, England, in 1962 and 1966, respectively.

After holding postdoctoral fellowship positions at Reading University and Bell telephone Laboratories, he joined GTE Laboratories in 1970 as a member of the technical staff. In 1972, he moved to McDonnell Douglas Research Laboratories, where he became a Research Staff Scientist. He joined the Georgia Tech Research Institute, Atlanta, GA, in 1981 and is currently a GTRI Fellow and Chief of the Quantum Microstructures Branch. His current research interests include optoelectronic properties of heterostructures, molecular beam epitaxy of II-IV and III-V semiconductors, and the growth and characterization of phosphor materials.



NON-PARABOLIC HYDRODYNAMIC FORMULATIONS FOR THE SIMULATION OF INHOMOGENEOUS SEMICONDUCTOR DEVICES

A. W. SMITH and K. F. BRENNAN

School of Electrical and Computer Engineering and Microelectronics Research Center,
Georgia Institute of Technology, Atlanta, GA 30332-0250, U.S.A.

(Received 4 April 1995; in revised form 15 March 1996)

Abstract—Hydrodynamic models are becoming prevalent design tools for small scale devices and other devices in which high energy effects can dominate transport. Most current hydrodynamic models use a parabolic band approximation to obtain fairly simple conservation equations. Interest in accounting for band structure effects in hydrodynamic device simulation has begun to grow since parabolic models cannot fully describe the transport in state of the art devices due to the distribution populating non-parabolic states within the band. This paper presents two different non-parabolic formulations of the hydrodynamic model suitable for the simulation of inhomogeneous semiconductor devices. The first formulation uses the Kane dispersion relationship $(\hbar k)^2/2m = W(1 + \alpha W)$. The second formulation makes use of a power law $\{(\hbar k)^2/2m = xW^y \}$ for the dispersion relation. Hydrodynamic models which use the first formulation rely on the binomial expansion to obtain moment equations with closed form coefficients. This limits the energy range over which the model is valid. The power law formulation readily produces closed form coefficients similar to those obtained using the parabolic band approximation. However, the fitting parameters (x, y) are only valid over a limited energy range. The physical significance of the band non-parabolicity is discussed as well as the advantages/disadvantages and approximations of the two non-parabolic models. A companion paper describes device simulations based on the three dispersion relationships; parabolic, Kane dispersion and power law dispersion. Copyright © 1996 Elsevier Science Ltd

NOMENCLATURE

W	carrier energy
h	Planck's constant divided by 2π
\mathbf{k}	reciprocal lattice vector
α	non-parabolicity factor under the Kane dispersion relation, positional dependent
m	carrier mass at the band edge, a constant but positional dependent
x, y	adjustable parameters for the power law dispersion relation, both positional dependent
$g(\mathbf{k})$	density of states in momentum space
W_c	conduction band edge
E_f	electron quasi-fermi level
K_B	Boltzmann's constant
T_e	electron temperature
F_n	Fermi integral of various order
\mathcal{F}_n	Fermi integral divided by gamma function
Γ	gamma function
N_c	effective density of states
η	reduced energy $\{(E_n - E_c)/(K T_e)\}$
E_g	semiconductor band gap
∇, ∇_r	gradient operator in physical space
∇_k	gradient operator in \mathbf{k} space
\mathbf{v}_g	group velocity
f	distribution function
f_0	equilibrium distribution function
τ	relaxation time
ϵ_c	conduction band potentials (electrostatic, affinity and band gap narrowing)
n	electron concentration
\mathbf{I}	identity matrix

$$g = \frac{1}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{3/2} \quad \Psi = 1 + 2\alpha W \quad Y = 1 + \alpha W$$

$$\Omega = (\mathcal{F}_{1/2} + \frac{1}{4}\alpha K T_e \mathcal{F}_{3/2})$$

$$A = (\mathcal{F}_{1/2} + \frac{1}{4}\alpha K T_e \mathcal{F}_{1/2})$$

1. INTRODUCTION

The use of hydrodynamic models for device simulation are becoming common as characteristic device dimensions continue to decrease. Hot electron effects can play a dominant role in carrier transport for high mobility semiconductors. Current hydrodynamic models consist of a set of conservation equations derived by taking moments of the Boltzmann transport equation (BTE). During the derivation of the conservation equations the parabolic band approximation is used to obtain rather simple coefficients on the forcing terms in the flux equations. By relying on the parabolic band approximation higher-order energy transport effects due to variation in the band structure are neglected. Interest in accounting for band structure effects in hydrodynamic device simulation has begun to grow because parabolic models cannot adequately account for high energy effects in semiconductors with

non-parabolic band structures. Non-parabolic band formulations have a history dating back to the 1950s[1-3]. However drift-diffusion models and more specifically hydrodynamic simulators with non-parabolic band formulations are a very recent topic of research.

Several non-parabolic hydrodynamic models have been reported for homogeneous material systems[4-7] using the Kane dispersion relationship[3]. The general functional form obtained is similar to parabolic hydrodynamic models with first-order corrections on the diffusion term. Azoff[8] derived a hydrodynamic model suitable for degenerate heterostructure semiconductors though the final form of the equations was not directly amenable to current device simulation codes. However, Azoff clearly showed that a forcing term due to a gradient in the non-parabolicity factor exists. Woolard *et al.*[9] presented a non-parabolic hydrodynamic model based on moments of the velocity and energy (u , $W(k)$) instead of the momentum and momentum squared ($\hbar k$, $\hbar^2 k^2$). This leads to a simpler energy conservation equation. However, the non-parabolic coefficient in the field term and the forcing terms due to non-uniform band structure were neglected in the other moment equations. Cassi and Riccò[10] introduced an alternative to the Kane relation in the form of a power law for the dispersion relationship. Instead of using the classical Kane dispersion law relating the energy and momentum, the band was fit over a specified energy range using two adjustable parameters. The approximations and assumptions implied by assuming the power law formulation were absent. It will be shown below that the power law formulation of the dispersion relation leads to a more simplistic and compact formulation than the classical Kane expression.

The purpose of this paper is to introduce two non-parabolic hydrodynamic formulations suitable for the simulation of devices with inhomogeneous material layers. The final form of the conservation equations will be in a form which will allow incorporation into existing device simulation codes, similar to the parabolic formulation. These different formulations are based on different choices for the dispersion relationship, one uses the standard Kane dispersion for non-parabolic bands and the second uses a power law relationship[10]. For comparison the parabolic hydrodynamic formulation will also be presented. The form of the conservation equations are strongly affected by the non-parabolicity factor of the bands, the choice of the dispersion relationship and the assumptions made to simplify the coefficients. As in the case of the parabolic formulation, both non-parabolic formulations require estimates of higher-order moments to provide mathematical closure of the relationships. It will be shown that more physical

insight can be obtained by examining the terms from the power law formulation due to their similarity to the parabolic formulation. The power law is advantageous in that the terms of the conservation equations are the same as in the parabolic formulation, except for a simple multiplicative constant when Boltzmann statistics are employed.

2. DISPERSION RELATIONS AND CARRIER CONCENTRATION

The two non-parabolic dispersion relations relating the energy to the momentum are:

$$W(1 + \alpha W) = \frac{(\hbar k)^2}{2m}, \quad xW^y = \frac{(\hbar k)^2}{2m}. \quad (1)$$

Where α is the non-parabolicity factor in the Kane dispersion relation. This parameter is usually calculated from a $k \cdot P$ perturbation approach to the band structure. The α factor can also be estimated in terms of known parameters of the semiconductor[10]. The second non-parabolic dispersion relation is the power law formulation of Cassi and Riccò[10] which is not a first-order approach to the band structure as compared to the Kane dispersion relation. Instead, the parameters x and y are obtained by a best fit to eqn (1a) over a specified energy range. In Ref. [10] this energy range was (1.5 eV, 3.0 eV) and the authors produced a very good fit for $\alpha = 0.4789$, determined by inference from x and y and is reproduced in Fig. 1. The formulae for the carrier concentration using these two dispersion relations and assuming Fermi-Dirac statistics are (the binomial expansion has been employed for eqn (2a)):

$$n = N_c(T_c) [\mathcal{F}_{1/2}(\eta) + \frac{15}{4} \alpha K_B T_c \mathcal{F}_{3/2}(\eta)]$$

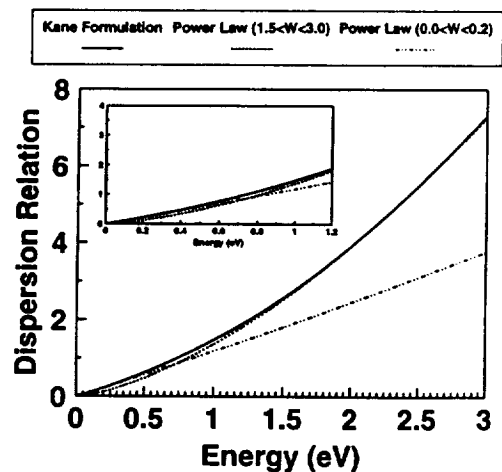


Fig. 1. Dispersion relations for the Kane model (solid line, $\alpha = 0.4789$), the power law formulation (dashed line, $x = 1.365$, $y = 1.52$ fit over high energy, $1.5 \leq W \leq 3.0$ eV) and the power law formulation (dotted line, $x = 1.185$, $y = 1.052$ fit over low energy, $0.0 \leq W \leq 0.2$ eV). The insert shows the detail of the low energy range.

g the terms
ue to their
The power
rms of the
as in the
mple multi-
statistics are

RRIER

ions relating

$$\frac{\hbar^2 k^2}{2m} \quad (1)$$

in the Kane
is usually
roach to the
estimated in
nductor[10].
lation is the
cò[10] which
structure as
on. Instead,
a best fit to
In Ref. [10]
l the authors
determined
ed in Fig. 1.
ration using
ning Fermi-
ion has been

$z(\eta)$



del (solid line,
dashed lined,
; $W \leq 3.0$ eV)
ie, $x = 1.185$,
(V). The insert
ange.

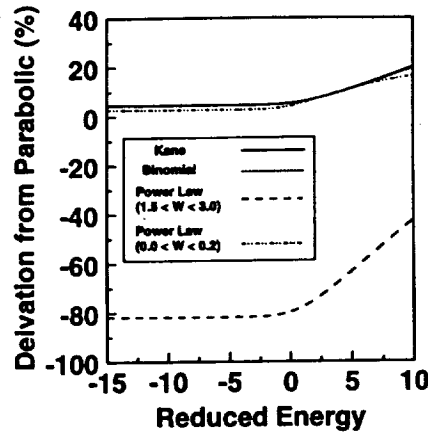


Fig. 2. Deviation, from the parabolic case, of the carrier concentration as a function of the reduced energy for the full α formulation (solid line, $\alpha = 0.4789$), binomial α (dotted), power law fit at high energy (dashed, $x = 1.365$, $y = 1.52$) and power law fit at low energy (dashed, $x = 1.185$, $y = 1.052$) dispersion relations as a function of the reduced energy (η) at $T_e = 300$ K.

$$n = N_c(T_e) x^{3/2} y (K_B T_e)^{(3/2)y-1} \left[\frac{\Gamma(\frac{3}{2}y)}{\Gamma(\frac{3}{2})} \mathcal{F}_{(3/2)-1}(\eta) \right]. \quad (2)$$

In the case of the Kane dispersion the non-parabolicity is a perturbation to the parabolic model regardless of the fact that Fermi-Dirac statistics were used in the derivation instead of Boltzmann statistics. In the power law case, unlike the α formulation, the carrier concentration will only differ from the parabolic by a multiplicative constant if Boltzmann statistics are assumed.

Within the hydrodynamic simulation of a device the carrier concentration is calculated by directly solving the continuity equations. Inspection of eqns (2a, b) shows that the concentration in both cases is a function of the reduced energy, η . The determination of η at each position is crucial to the successful simulation of a device since most of the factors within the transport equations i.e. number of ionized dopants, Shockley-Read-Hall recombination etc., are functions of η as well. Therefore, from the knowledge of the carrier concentration at any specific position the reduced energy, η , is calculated by suitably inverting eqns (2a, b). With these facts in mind it is illustrative to compare the non-parabolic formulations for the carrier concentration to the parabolic formulation at various reduced energy values. Figure 2 displays the deviation from the parabolic formulation for $\alpha = 0.4789$, $x = 1.365$ and $y = 1.52$ [10]. The case of the full Kane dispersion relation before binomial expansion is also given in Fig. 2 and shows that the binomial expansion is justified for this case of the non-parabolicity parameter; the two curves are almost identical throughout the entire reduced energy range. Notice also that the deviation of the alpha non-parabolicity

formulations is at most 5% from the parabolic case until the reduced energy exceeds zero, when the difference in the orders of the Fermi integral is larger. However, in the case of the power law fit at high energy, the deviation from the parabolic case is more than 82% and more than 87% from the α formulation. To explain this large difference the plots of the dispersion relations must be re-examined, Fig. 1. At the high energy part of the curve the power law and α formulations have nearly identical dispersion relations. The insert in Fig. 1 shows the low energy range and shows the maximum deviation of the two curves. Though there is a greater deviation between the two curves at lower energy than at high energy, the maximum difference between the two curves is insufficient by itself to fully account for the large difference in the carrier concentrations. However, carrier concentration depends upon both the density of states and the distribution function. Since the distribution function weights more heavily to lower energy states, a mismatch at low energy is strongly reflected in the product of the distribution function times the density of states and subsequently the carrier concentration. Therefore, in order to obtain a matched value of the carrier concentration, it is critical to have a closer fit to the low energy range of the band. For comparison, a low energy fit was made using a different set of values for x and y , $x = 1.185$ and $y = 1.052$. These values of x and y were determined by fitting the power law dispersion relation over the low energy range ($0.0 \text{ eV} \leq W \leq 0.2 \text{ eV}$). As can be seen from the inset of Fig. 1, the dispersion relations match very closely at low energy but deviate significantly at higher energy. Figure 2 also shows the calculated carrier concentrations for various reduced energy values using the low energy power law fit factors. Comparing the power law fit to the high energy range and applying the power law fit to the low energy range of the band produces concentration values which closely match those obtained using the α formulation.

3. PARTICLE FLUX EQUATIONS

The general flux ($\Theta = v_x$) conservation equation is determined by taking moments of the BTE as

$$\frac{\partial}{\partial t} (\bar{v}_x f) + \nabla \cdot (f \bar{v}_x \bar{v}_x) - f \frac{\hbar^2 (\mathbf{k} \cdot \nabla)}{m(k)} \cdot \nabla \left(\frac{1}{m(W)} \right) - f \bar{F} \cdot \frac{1}{m(k)} \mathbf{I} + f \bar{F} \cdot \frac{3(\mathbf{k} \cdot \nabla) m(k)}{m(k)^2} = \bar{v}_x \frac{-(f - f_0)}{\tau}. \quad (3)$$

Notice that the third, fourth and fifth terms on the left-hand-side are tensor products. The factor of 3 in the fifth term is due to the order of parenthesis in the original moment equation. This moment equation cannot be processed further until some functional

form of the effective mass is assumed, which depends on the choice of the dispersion relation. Using the parabolic dispersion and the two dispersion relations in eqn (1), the resulting flux equations are (positional subscript on the gradient operators has been removed):

$$\frac{\partial}{\partial t}(\bar{v}_x f) + \frac{2}{3m_e} \nabla(fW) + \nabla \epsilon_c \frac{f}{m_e} - \frac{\nabla m_e}{m_e^2} fW = \bar{v}_x \frac{-(f-f_0)}{\tau}, \quad (4)$$

$$\begin{aligned} \frac{\partial}{\partial t}(\bar{v}_x f) + \frac{2}{3m_e} \nabla \left(f \frac{WY}{\Psi^2} \right) - \left(f [1 + \frac{4}{3} \alpha WY] \frac{WY}{\Psi^4} \right) \frac{\nabla m_e}{m_e^2} \\ + \left(f \frac{W^2(-1 + \frac{4}{3} Y^2)}{m_e \Psi^4} \right) \nabla \alpha + \left(\frac{f}{m_e \Psi^3} \right) \nabla \epsilon_c \\ = \bar{v}_x \frac{-(f-f_0)}{\tau}, \end{aligned} \quad (5)$$

$$\begin{aligned} \frac{\partial}{\partial t}(\bar{v}_x f) + \frac{2}{3m_e x y^2} \nabla(fW^{2-y}) \\ + \left[\frac{\nabla m_e}{m_e} + \frac{\nabla x}{x} \right] \frac{fW^{2-y}(y-4)}{3m_e x y^3} \\ + \left(\frac{fW^{1-y}}{m_e x y} \right) \left(\frac{2}{y} - 1 \right) \nabla \epsilon_c \\ + \left(\frac{fW^{2-y}}{m_e x y^3} \right) \left(-\frac{1}{3}(1 + 2 \ln(W)) + y \ln(W) \right) \nabla y \\ = \bar{v}_x \frac{-(f-f_0)}{\tau}. \end{aligned} \quad (6)$$

Each of these equations must be integrated over all k space or equivalently over energy using the density of states[11]. Before the integration is performed one more assumption must be made, that the relaxation time is independent of k or W . If the constant relaxation time assumption is not made then the energy dependence of the relaxation time must be moved through the gradient operator on the second term in each of equations (4)–(6) and a term accounting for the gradient of the relaxation time must be re-created. Changing the integration from k space to energy space, substituting for the mobility, ($\mu = \tau/m$) and making $v_x f_0$ equal to zero gives the following integral equations:

$$\begin{aligned} \tau \frac{\partial(nv)}{\partial t} + \frac{2\mu}{3} \nabla \left[g \int fW^{3/2} dW \right] - \mu \frac{\nabla m_e}{m_e} g \int fW^{3/2} dW \\ + \mu \nabla \epsilon_c g \int fW^{1/2} dW = -nv, \end{aligned} \quad (7)$$

$$\begin{aligned} \tau \frac{\partial(nv)}{\partial t} + \frac{2\mu}{3} \nabla \left[g \int f \frac{(WY)^{3/2}}{\Psi^2} dW \right] \\ - \mu \frac{\nabla m_e}{m_e} g \int f \frac{(WY)^{3/2}(1 + \frac{4}{3} \alpha WY)}{\Psi^3} dW \\ + \mu \nabla \alpha g \int f \frac{W^2(-1 + \frac{4}{3} Y^2)(WY)^{1/2}}{\Psi^3} dW \\ + \mu \nabla \epsilon_c g \int f \frac{(WY)^{1/2}}{\Psi^2} dW = -nv, \end{aligned} \quad (8)$$

$$\begin{aligned} \tau \frac{\partial(nv)}{\partial t} + \frac{2\mu\sqrt{x}}{3y} \nabla g \int fW^{(y/2)+1} dW \\ + \frac{\mu\sqrt{x}\nabla m_e}{3y^2 m_e} [y-4]g \int fW^{(y/2)+1} dW \\ + \frac{\mu\sqrt{x}\nabla x}{3xy^2} [y-4]g \int fW^{(y/2)+1} dW \\ + \frac{\mu\sqrt{x}\nabla y}{y^3} g \int fW^{(y/2)+1} [-2-4 \ln(W) \\ + 3y \ln(W)] dW \\ + \mu\sqrt{x}\nabla \epsilon_c \left(\frac{2}{y} - 1 \right) g \int fW^{y/2} dW = -nv. \end{aligned} \quad (9)$$

To produce closed form solutions for the integrals in eqn (8) the binomial expansion is used repeatedly and all terms of order α^2 or higher are set to zero, to finally become

$$\begin{aligned} \tau \frac{\partial(nv)}{\partial t} + \frac{2\mu}{3} \nabla \left[g \int fW^{3/2} \left(1 - \frac{\alpha W}{2} \right) dW \right] \\ - \mu \frac{\nabla m_e}{m_e} g \int fW^{3/2} \left(1 - \frac{19\alpha W}{6} \right) dW \\ + \mu \nabla \alpha g \int fW^{3/2} \left(\frac{1}{3} + \frac{5\alpha W}{6} \right) dW \\ + \mu \nabla \epsilon_c g \int fW^{1/2} \left(1 - \frac{7\alpha W}{2} \right) dW = -nv. \end{aligned} \quad (10)$$

The only term in the power law flux equation (eqn (9)) which will require an expansion is the ∇y term due to the $\ln(W)$ factor in the integrand. The expansion may only be required for certain choices of the distribution, but to maintain

Table 1. Terms (prefactors) within the integrands of the forcing coefficients which are graphically compared

Formulation	Equation number	Diffusion term	Mass term ∇m	Field term $\nabla \epsilon_c$
parabolic	(7)	$W^{3/2}$	$W^{3/2}$	$W^{3/2}$
α formulation	(8)	$\frac{W(1 + \alpha W)^{1/2}}{(1 + 2\alpha W)}$	$\frac{W(1 + \alpha W)^{1/2}(1 + 4\alpha W)}{(1 + 2\alpha W)^3}$	$\frac{W^{3/2}(1 + \alpha W)^{1/2}}{(1 + 2\alpha W)^3}$
α with binomial expansion used	(10)	$W^{3/2}\left(1 - \frac{\alpha W}{2}\right)$	$W^{3/2}\left(1 - \frac{19\alpha W}{6}\right)$	$W^{3/2}\left(1 - \frac{7\alpha W}{2}\right)$
power law	(12)	$\frac{x^{1/2}}{y} W^{3/2 + (v/2)}$	$\frac{x^{1/2}}{3y^3} [4 - y] W^{3/2 + (v/2)}$	$\frac{x^{1/2}}{y} \frac{(2 - y)}{y} W^{3/2}$

generality it is applied for all distributions. A parabolic interpolation, using the points, 0, 0.5 and 1, to the part of the equation containing the log term is:

$$\int W^{(v/2)+1} \ln(W) dW \approx 4(0.5)^{(v/2)+1} \ln(0.5) \times \int f(W^2 - W) dW. \quad (11)$$

With the above substitution the power law flux equation can be written as:

$$\begin{aligned} \tau \frac{\partial(nv)}{\partial t} + \frac{2\mu\sqrt{x}}{3y} \nabla \left[g \int W^{(v/2)+1} dW \right] \\ + \frac{\mu\sqrt{x}\nabla m_c}{3y^2 m_c} [y - 4] g \int W^{(v/2)+1} dW \\ + \frac{\mu\sqrt{x}\nabla x}{3xy^2} [y - 4] g \int W^{(v/2)+1} dW \\ - \frac{2\mu\sqrt{x}\nabla y}{y^3} g \int W^{(v/2)+1} dW \\ + \frac{\mu\sqrt{x}\nabla y}{y^3} g [-4 + 3y] 4 \ln(0.5) (0.5)^{(v/2)+1} \\ \times \int f(W^2 - W) dW + \mu\sqrt{x}\nabla \epsilon_c \left(\frac{2}{y} - 1 \right) g \\ \times \int W^{3/2} dW = -nv. \end{aligned} \quad (12)$$

The approximation in eqn (11) tends to degrade as y increases. However, this forcing term will not appear in simulations of homogeneous materials. In the case of inhomogeneous materials systems this term may be quite small as y has only a limited range, ($1 \leq y \leq 2$).

One of the goals of this paper is to derive a hydrodynamic model suitable for fast efficient

simulation of state of the art devices. As previously stated the parabolic model is inadequate for certain material systems due to the fact that their band structures deviate dramatically from a parabolic shape at high energies. In these situations, a non-parabolic band structure provides a more accurate description. As discussed above, the full α formulation captures the desired physics, but is unattractive because the coefficients within the transport equations need to be numerically evaluated. In contrast, the binomial α formulation and the power law formulation provide closed form expressions for the transport coefficients. However, the question remains as to their ability to display the proper trends; how closely do the coefficients match the full α formulation? We will show that the binomial α formulation has a limited energy and non-parabolicity range due to the binomial approximation. Non-physical results will be obtained if the formulation is extended into regions outside the binomial limit. The power law does not produce non-physical results but more closely matches the parabolic formulation when fit to the low energy part of the band.

The coefficients on similar forcing terms in the various formulations can now be compared to examine the impact of the non-parabolicity factor and approximations made during the derivation. The comparison is done at this time to avoid any confusion from assuming a distribution function. Table 1 lists the forcing terms which are compared and the terms within the integrands involved. Table 2 contains the values of the non-parabolicity factors at which these factors are compared, this includes values of α from 0.04 to 4.0 and appropriate (x, y) values fit to these α values over two different energy ranges

Table 2. Values of the non-parabolicity factors which are compared. This includes α values from 0.04 to 4.0 and (x, y) values fit to these α values over two different energy ranges (0.0, 0.2) and (1.5, 3.0)

Case	α	x, y {0.0, 0.2}	x, y {1.5, 3.0}
a	0.04	1.0148, 1.0045	1.0207, 1.0827
b	0.4789	1.185, 1.052	1.365, 1.52
c	4.0	2.975, 1.322	4.8233, 1.901

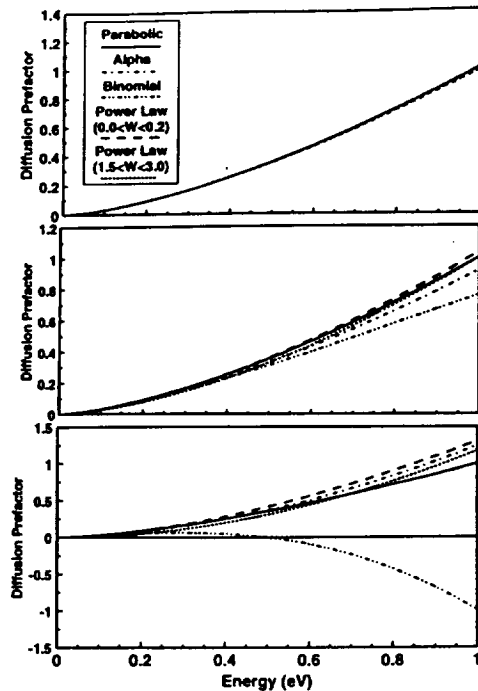


Fig. 3. Comparison of the integrand coefficients (prefactors) occurring on the diffusion term of the particle flux formulations. See Tables 1 and 2 for the exact form of the equations and degrees of non-parabolicity considered.

(0.0, 0.2) and (1.5, 3.0). Figure 3 shows the prefactor for the diffusion term from all the flux equations, Fig. 3(a) is for slightly non-parabolic bands ($\alpha = 0.04$); Fig. 3(b) is for the non-parabolicity in Ref. [10] ($\alpha = 0.4789$) and 3(c) is for a highly non-parabolic band ($\alpha = 4.0$). Figures 4(a-c) and 5(a-c) display the prefactors on the forcing terms from changes in the effective mass and the field for the three degrees of non-parabolicity. For all three forcing terms at the lowest values of non-parabolicity (Figs. 3a, 4a, 5a) the prefactors compare favorably with the parabolic formulation. For the diffusion term (Fig. 3(a-c)) as the non-parabolicity factor increases the two cases of power law formulation match very closely to the full α formulation. On the other hand the α formulation that utilizes the binomial expansion is very different, especially at the highest value of non-parabolicity considered. From Fig. 3(c) it is clear that the binomial α formulation is clearly incorrect if the energy exceeds 0.5 eV. Even before this point the diffusion will be underestimated. In the case of the mass term, Fig. 4, the power law formulation which is fit over the low energy range is much closer to the parabolic case, as expected due to the small change in the fitting parameters from their parabolic values. However, the power law with the parameters fit over a larger energy range more closely matches the full α formulation. As in the case of the diffusion term, the binomial α formulation severely underestimates the effect of this forcing term

especially as the non-parabolicity is increased and is limited to energies less than 0.1 eV for $\alpha = 4.0$. The coefficients for the field term, Fig. 5, follow the same conclusions as for the mass term. From these figures it is clear that the binomial α formulation has a very limited energy range of validity as the non-parabolicity factor is increased. Using this formulation at higher energies or high non-parabolicity factors can give un-physical results due to the prefactors changing sign. On the other hand, the power law formulation with parameters fit over a small energy range will tend to produce results which more closely match the parabolic band model. It will not produce un-physical trends and does appear to have a larger range of validity for both energy and non-parabolicity factors. The case of the power law with parameters fit over a large energy range more closely matches the full α formulation in terms of the forcing coefficients. However, due to the problems previously described for the calculation of the carrier concentration this advantage may be immaterial.

Table 1 can also be used to gain some physical insights into transport in the non-parabolic band structures, especially using the power law formulation. First notice that as the non-parabolicity factor is decreased ($\alpha \rightarrow 0$, $x \rightarrow 1$, $y \rightarrow 1$) all the coefficients reduce to the parabolic case indicating that all three formulations are equivalent in this respect. As the non-parabolicity factor and energy increases the binomial α formulation can actually predict a change in the sign of a forcing term. In the full α formulation

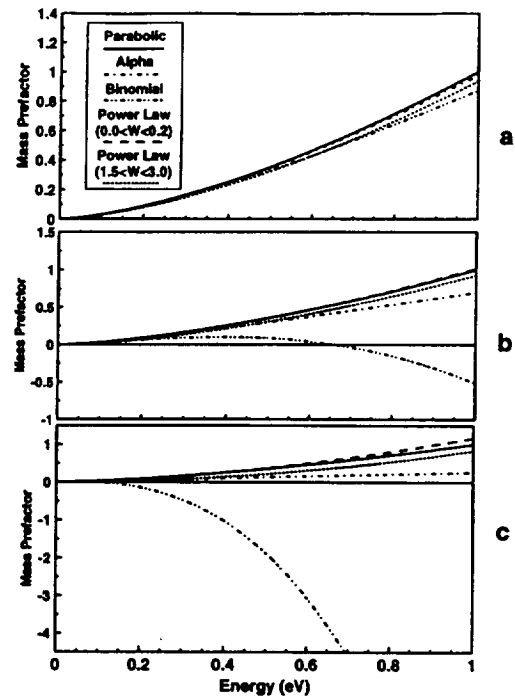


Fig. 4. Comparison of the integrand coefficients (prefactors) occurring on the gradient of the effective mass term of the particle flux formulations. See Tables 1 and 2 for the exact form of the equations and degrees of non-parabolicity considered.

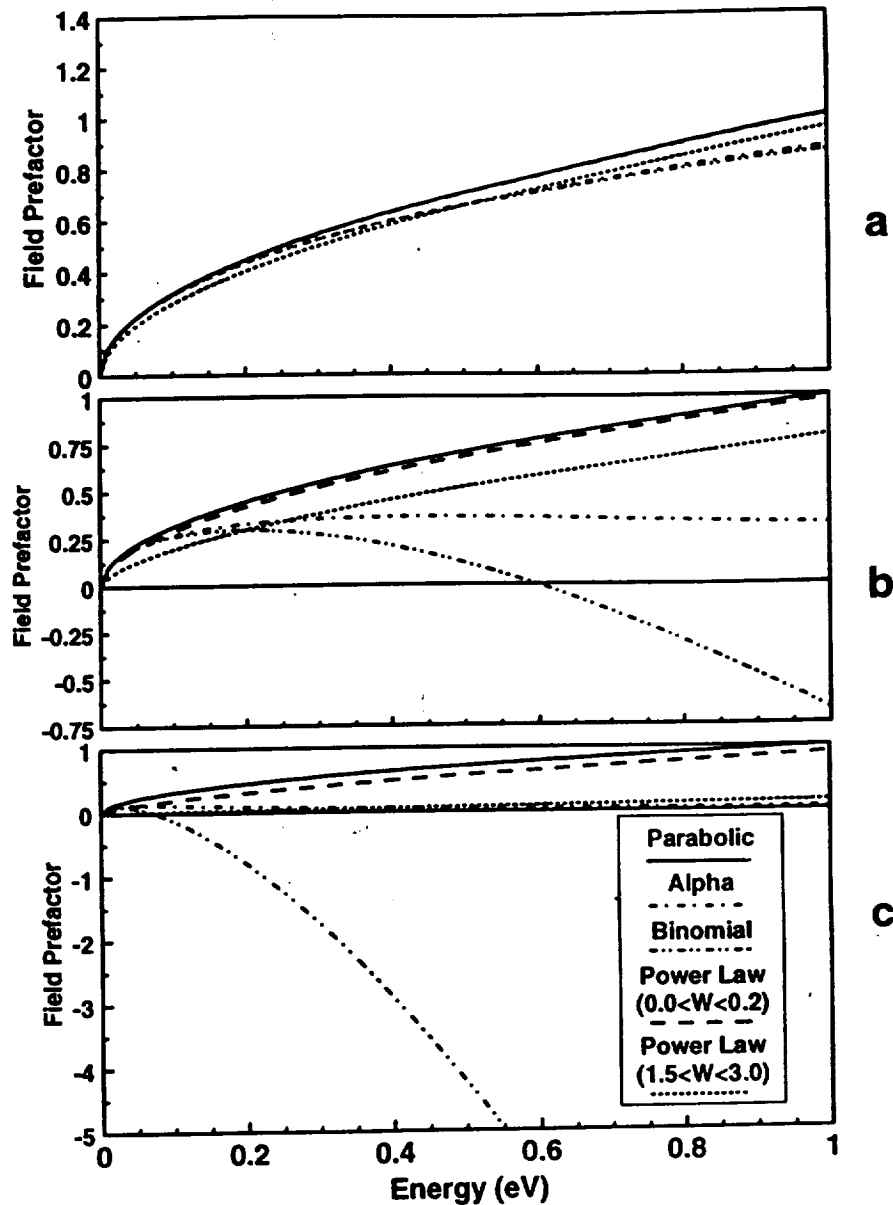


Fig. 5. Comparison of the integrand coefficients (prefactors) occurring on the gradient of the potential term occurring in the particle flux formulations. See Tables 1 and 2 for exact form of the equations and degrees of non-parabolicity considered.

the sign on the forcing terms does not change but it is unclear as to how the term decreases. In the case of the power law, when the parameter γ equals 2 it is obvious that the field term will become identically zero. The only terms which will be non-zero in the flux equation will be the diffusion, gradient in mass and gradient in the non-parabolicity terms. This can be explained with the use of the power law energy equation {eqn (1b)} and the group velocity equation. When γ equals 2 there is a linear relationship between the momentum and energy, the bands are V shaped. The group velocity is proportional to the gradient of the energy with respect to k , which for the case of γ

equals 2 means that the group velocity is a constant. Therefore, no matter how much force is applied to the electron by external forces the velocity is not increased. The only factors which can produce current are gradients which can change this fixed velocity, i.e. changes in mass or non-parabolicity factor, or changes in the number of carriers moving at this fixed velocity, diffusion. Therefore non-parabolic formulations which do not include coefficients to diminish the field term as the non-parabolicity increases overestimate the flux.

At this point in the derivation a recursion relation must be formulated, a distribution function assumed,

or some other mathematical method (minimum-maximum theorem) must be used to provide mathematical closure for eqns (7), (10) and (12). Since, a goal of this paper is the formulation of models suitable for the numerical simulation of devices the first two options are explored[12]. A recursion relation would allow moments of higher-order to be approximated by lower-order moments, the lower-orders are calculated from the conservation equations. This option does not require that a specific form of the distribution function be used, unless the recursion relations are based on a specific distribution. However, in the case of non-parabolic bands the standard recursion relations may no longer be applicable[9]. Therefore, this option was not pursued. The other option and the one chosen for this work, is to assume a specific form for the distribution function, higher moments can then be calculated based on the known distribution function. Some of the choices for the distribution function include heated Maxwellian, shifted and heated Maxwellian, heated Fermi-Dirac, or shifted and heated Fermi-Dirac. Since the Maxwellian distributions can be recovered by relaxing the degeneracy, the Fermi-Dirac distributions were the only ones considered for this work.

In the non-parabolic formulation a simple relation between energy and velocity will not exist due to the change in the density of states. Also higher-order powers of the energy are required to close the relationships in the α formulation, this will require cross product terms involving the temperature and the velocity. In addition, the power law formulation has non-integer powers of the energy which will be very difficult to evaluate for the shifted and heated distributions. Due to these conditions and the fact that all the formulations break down as the energy rises, the heated Fermi-Dirac distribution was used to close the relationships. The flux equation in the binomial α formulation is:

$$\begin{aligned}
 -n\bar{v} &= \mu K T_e \\
 &\times \left[\frac{\mathcal{F}_{1/2}^2 + \frac{10}{4}\alpha K T_e \mathcal{F}_{3/2} \mathcal{F}_{1/2}}{\mathcal{F}_{1/2} \mathcal{F}_{-1/2} + \frac{15}{4}\alpha K T_e \left(\mathcal{F}_{1/2}^2 + \frac{\mathcal{F}_{-1/2} \mathcal{F}_{3/2}}{2} \right)} \right] \nabla n \\
 &+ \mu n \left[\frac{\mathcal{F}_{1/2} - \frac{21}{4}\alpha K T_e \mathcal{F}_{3/2}}{\Omega} \right] \nabla \epsilon_c \\
 &- \frac{\mu n K T_e}{\Lambda} \\
 &\times \left[\mathcal{F}_{1/2} + \frac{\mathcal{F}_{1/2}^2 + 20\alpha K T_e \mathcal{F}_{3/2} \mathcal{F}_{-1/2}}{2\Omega} \right] \frac{\nabla m}{m} \\
 &+ \frac{\mu n K T_e}{\Lambda \Omega}
 \end{aligned}$$

$$\begin{aligned}
 &\times \left[\frac{5}{2} \mathcal{F}_{-1/2} \mathcal{F}_{3/2} - \mathcal{F}_{1/2}^2 - \frac{\mathcal{F}_{1/2}^3}{2\Omega} \right] \frac{\nabla T_e}{T_e} \\
 &- \frac{4\mu n (K T_e)^2}{\Omega} \\
 &\times \left[\frac{10\mathcal{F}_{1/2} \mathcal{F}_{3/2}}{\Lambda} - \frac{175}{4} \alpha K T_e \mathcal{F}_{1/2} + \frac{\mathcal{F}_{1/2}^2 \mathcal{F}_{3/2}}{\Lambda \Omega} \right] \nabla \alpha.
 \end{aligned} \tag{13}$$

The flux equation in the power law formulation becomes:

$$\begin{aligned}
 -n\bar{v} &= \frac{2\mu}{3xy^2} (K T_e)^{2-y} \frac{\left(\frac{y}{2} + 1\right) \Gamma\left(\frac{y}{2} + 1\right) \mathcal{F}_{y/2}}{\left(\frac{3}{2}y - 1\right) \Gamma\left(\frac{3}{2}y - 1\right) \mathcal{F}_{(3y/2) - 2}} \nabla n \\
 &+ \frac{(2-y)\mu n}{xy^2} (K T_e)^{1-y} \frac{\Gamma\left(\frac{y}{2} + 1\right) \mathcal{F}_{y/2}}{\Gamma\left(\frac{3}{2}y\right) \mathcal{F}_{(3y/2) - 1}} \nabla \epsilon_c \\
 &+ \frac{\mu n}{xy^2} (K T_e)^{2-y} \left[\frac{-y\left(\frac{y}{2} + 1\right) \Gamma\left(\frac{y}{2} + 1\right) \mathcal{F}_{y/2}}{\left(\frac{3}{2}y - 1\right) \Gamma\left(\frac{3}{2}y - 1\right) \mathcal{F}_{(3y/2) - 2}} \right. \\
 &\quad \left. + \frac{(4+y)\Gamma\left(\frac{y}{2} + 2\right) \mathcal{F}_{(y/2) + 1}}{3\Gamma\left(\frac{3}{2}y\right) \mathcal{F}_{(3y/2) - 1}} \right] \frac{\nabla T_e}{T_e} \\
 &- \frac{\mu n}{xy^2} (K T_e)^{2-y} \left[\frac{\left(\frac{y}{2} + 1\right) \Gamma\left(\frac{y}{2} + 1\right) \mathcal{F}_{y/2}}{\left(\frac{3}{2}y - 1\right) \Gamma\left(\frac{3}{2}y - 1\right) \mathcal{F}_{(3y/2) - 2}} \right. \\
 &\quad \left. - \frac{4\Gamma\left(\frac{y}{2} + 2\right) \mathcal{F}_{(y/2) + 1}}{3\Gamma\left(\frac{3}{2}y\right) \mathcal{F}_{(3y/2) - 1}} \left(1 - \frac{1}{y}\right) \right] \left(\frac{\nabla m}{m} + \frac{\nabla x}{x} \right)
 \end{aligned}$$

$$\begin{aligned}
& + \frac{\mu n}{xy^3} (KT_e)^{2-y} \left[\frac{-2\left(\frac{y}{2}+1\right)\Gamma\left(\frac{y}{2}+1\right)\mathcal{F}_{y/2}}{3\left(\frac{3}{2}y-1\right)\Gamma\left(\frac{3}{2}y-1\right)\mathcal{F}_{(3y/2)-2}} \right. \\
& \times \left(1 + \frac{3}{2}y \ln(KT_e) \right. \\
& + \frac{\Gamma\left(\frac{y}{2}+2\right)\mathcal{F}_{(y/2)+1}}{3\Gamma\left(\frac{3}{2}y\right)\mathcal{F}_{(3y/2)-1}} (-4+y \ln(KT_e)) \\
& + (3y-4)4 \ln(0.5)(0.5)^{(y/2)+1} (KT_e)^{y/2} \\
& \left. \left. \times \left(\frac{KT_e \Gamma(3)\mathcal{F}_2 - \Gamma(2)\mathcal{F}_1}{\Gamma\left(\frac{3}{2}y\right)\mathcal{F}_{(3y/2)-1}} \right) \right] \nabla y. \quad (14)
\end{aligned}$$

The reader can verify that as the non-parabolicity factors are diminished, ($\alpha \rightarrow 0$, $x, y \rightarrow 1$), the two non-parabolic formulations reduce to the standard parabolic case. The equations also reduce to simpler forms when the degeneracy effects are ignored (all orders of Fermi integral reduce to exponentials). The flux equations can be discretized using normal techniques to produce comparable simulation codes under various assumptions[13]

4. ENERGY FLUX EQUATIONS

For the energy flux equation ($\Theta = Wv_x$) the general conservation equation is

$$\begin{aligned}
& \frac{\partial}{\partial t} (W \bar{v}_x) + \nabla \cdot (f W \bar{v}_x \bar{v}_x) - f \bar{v}_x \bar{v}_x \cdot \nabla W \\
& - f W \frac{\hbar^2 (\mathbf{k}\mathbf{k})}{m(k)} \cdot \nabla \left(\frac{1}{m(W)} \right) - f \mathbf{F} \cdot \bar{v}_x \bar{v}_x - f \mathbf{F} \\
& \cdot \frac{W}{m(k)} \mathbf{I} + f \mathbf{F} \cdot \frac{3W(\mathbf{k}\nabla_k m(k))}{m(k)^2} \\
& = W \bar{v}_x \frac{-(f-f_0)}{\tau}. \quad (15)
\end{aligned}$$

Notice that the fourth through seventh terms on the left-hand-side are tensor products. As in the case of the particle flux moment the energy flux moment equation cannot be processed further until some functional form of the effective mass is assumed. By making similar assumptions, substitutions and approximations {binomial expansion, Fermi-Dirac statistics, equipartition of energy . . .} the energy flux equations for the three dispersion relationships become

$$\begin{aligned}
& -S = \frac{1}{2} \mu n K T_e \left[K T_e \left[\frac{\mathcal{F}_{3/2}}{\mathcal{F}_{-1/2}} \right] \frac{\nabla n}{n} + \left[\frac{\mathcal{F}_{3/2}}{\mathcal{F}_{-1/2}} \right] \nabla \epsilon_c \right. \\
& - \frac{3K T_e}{2} \left[\frac{\mathcal{F}_{3/2}}{\mathcal{F}_{-1/2}} \right] \frac{\nabla m}{m} + \frac{1}{2} K T_e \frac{\mathcal{F}_{5/2}}{\mathcal{F}_{1/2}} \\
& \times \left[1 - \frac{3}{7} \frac{\mathcal{F}_{1/2} \mathcal{F}_{3/2}}{\mathcal{F}_{-1/2} \mathcal{F}_{5/2}} \right] \frac{\nabla T_e}{T_e} \left. \right]. \quad (16)
\end{aligned}$$

$$\begin{aligned}
& -S = \frac{5\mu(KT_e)^2}{2} \left[\frac{\mathcal{F}_{3/2} \Omega - \frac{1}{4} \alpha K T_e \mathcal{F}_{1/2} \mathcal{F}_{5/2}}{\Lambda \Omega} \right] \nabla n \\
& + \frac{5\mu n K T_e}{2} \left[\frac{\mathcal{F}_{3/2} - \frac{21}{4} \alpha K T_e \mathcal{F}_{5/2}}{\Omega} \right] \nabla \epsilon_c \\
& - \frac{5\mu n (K T_e)^2}{2} \left[\frac{3\mathcal{F}_{3/2}}{2\Lambda} - \frac{14\alpha K T_e \mathcal{F}_{1/2}}{\Omega} \right. \\
& \left. - \frac{21\alpha K T_e \mathcal{F}_{5/2} \mathcal{F}_{1/2}}{8\Lambda \Omega} \right] \frac{\nabla m}{m} \\
& + \frac{5\mu n (K T_e)^2}{2\Omega} \left[\frac{5}{2} \mathcal{F}_{5/2} - \frac{21}{8} \alpha K T_e \mathcal{F}_{7/2} \right. \\
& + \frac{\mathcal{F}_{1/2} \mathcal{F}_{5/2} - \frac{3}{20} \alpha K T_e \mathcal{F}_{1/2} \mathcal{F}_{7/2}}{\Omega} \\
& \left. - \frac{3\mathcal{F}_{3/2} \Omega - \frac{101}{4} \alpha K T_e \mathcal{F}_{1/2} \mathcal{F}_{5/2}}{2\Lambda} \right] \frac{\nabla T_e}{T_e} \\
& - \frac{5\mu n (K T_e)^3}{2\Omega} \left[\frac{3\mathcal{F}_{3/2}^2}{2\Lambda} - \frac{315}{16} \alpha K T_e \mathcal{F}_{9/2} \right. \\
& \left. + \frac{105\alpha K T_e \mathcal{F}_{1/2} \mathcal{F}_{3/2} \mathcal{F}_{5/2}}{16\Lambda \Omega} \right] \nabla \alpha. \quad (17)
\end{aligned}$$

$$\begin{aligned}
& -S = \frac{2\mu}{3xy^2} (KT_e)^{3-y} \\
& \times \frac{\left(\frac{y}{2}+2 \right) \Gamma\left(\frac{y}{2}+2 \right) \mathcal{F}_{(y/2)+1}}{\left(\frac{3}{2}y-1 \right) \Gamma\left(\frac{3}{2}y-1 \right) \mathcal{F}_{(3y/2)-2}} \nabla n \\
& + \frac{(8-3y)\mu n}{3xy^2} (KT_e)^{2-y} \frac{\Gamma\left(\frac{y}{2}+2 \right) \mathcal{F}_{(y/2)+1}}{\Gamma\left(\frac{3}{2}y \right) \mathcal{F}_{(3y/2)-1}} \nabla \epsilon_c \\
& + \frac{\mu n}{xy^2} (KT_e)^{3-y} \left[\frac{-y\left(\frac{y}{2}+2 \right) \Gamma\left(\frac{y}{2}+2 \right) \mathcal{F}_{(y/2)+1}}{\left(\frac{3}{2}y-1 \right) \Gamma\left(\frac{3}{2}y-1 \right) \mathcal{F}_{(3y/2)-2}} \right.
\end{aligned}$$

$$\begin{aligned}
& + \frac{(6+y)\Gamma\left(\frac{y}{2}+3\right)\mathcal{F}_{(y/2)+2}}{3\Gamma\left(\frac{3}{2}y\right)\mathcal{F}_{(3y/2)-1}} \left[\frac{\nabla T_e}{T_e} \right. \\
& - \frac{\mu n}{xy^2} (KT_e)^{3-y} \left[\frac{\left(\frac{y}{2}+2\right)\Gamma\left(\frac{y}{2}+2\right)\mathcal{F}_{(y/2)+1}}{\left(\frac{3}{2}y-1\right)\Gamma\left(\frac{3}{2}y-1\right)\mathcal{F}_{(3y/2)-2}} \right. \\
& - \frac{4\Gamma\left(\frac{y}{2}+3\right)\mathcal{F}_{(y/2)+2}}{3\Gamma\left(\frac{3}{2}y\right)\mathcal{F}_{(3y/2)-1}} \left(1-\frac{1}{y}\right) \left. \left(\frac{\nabla m}{m} + \frac{\nabla x}{x} \right) \right. \\
& + \frac{\mu n}{xy^3} (KT_e)^{3-y} \left[\frac{-2\left(\frac{y}{2}+2\right)\Gamma\left(\frac{y}{2}+2\right)\mathcal{F}_{(y/2)+1}}{3\left(\frac{3}{2}y-1\right)\Gamma\left(\frac{3}{2}y-1\right)\mathcal{F}_{(3y/2)-2}} \right. \\
& \times \left(1 + \frac{3}{2}y \ln(KT_e) \right) \\
& + \frac{\Gamma\left(\frac{y}{2}+3\right)\mathcal{F}_{(y/2)+2}}{3\Gamma\left(\frac{3}{2}y\right)\mathcal{F}_{(3y/2)-1}} (-4+y \ln(KT_e)) \\
& + (3y-4)4 \ln(0.5)(0.5)^{(y/2)+2} (KT_e)^{-(y/2)+1} \\
& \times \left[\frac{KT_e \Gamma(3)\mathcal{F}_2 - \Gamma(2)\mathcal{F}_1}{\Gamma\left(\frac{3}{2}y\right)\mathcal{F}_{(3y/2)-1}} \right] \nabla y. \quad (18)
\end{aligned}$$

First, eqn (16) corrects a sign error on the order of a Fermi integral which occurred in an earlier publication[13]. Again, the reader can verify that the non-parabolic formulations reduce to the parabolic case as the non-parabolicity factors are decreased. As in the case of the particle flux equations the energy flux equations can be discretized using normal techniques[13].

5. CONCLUSIONS

Two formulations of the hydrodynamic model have been presented for the simulation of non-parabolic inhomogeneous material systems, the standard Kane formulation and the power law formulation of Cassi and Riccò. Both forms reduce to the parabolic hydrodynamic model as the non-parabolicity factors are diminished. The forcing terms in the particle and energy flux equations due to

variations in the non-parabolicity factors of the bands as suggested by Azoff have been re-created in both non-parabolic formulations. It was shown that the binomial α formulation is suitable for the calculation of the carrier concentration but has a limited energy and non-parabolicity range when applied to the coefficients of the flux equations. Extending the binomial α formulation past these limits leads to non-physical terms in both the particle and energy flux equations. In the case of the power law formulation it was shown that when the adjustable parameters were fit to the high energy range the deviation in the carrier concentration from both the parabolic and α cases was extreme. When fit to the lower energy range the power law produced carrier concentrations comparable to the α formulations. The energy and non-parabolicity range of the power law formulation for the particle and energy flux equations is larger than the binomial α formulation, but it is still limited by the adjustable parameters. However, unlike the binomial α formulation the power law will not lead to physically unrealistic results, but will tend to more closely match the parabolic formulation when the adjustable parameters are fit over the low energy portion of the band. It is shown and argued using the power law formulation and a high non-parabolicity factor ($y=2$) that a dampening factor must exist on the field term in the particle flux equation to account for the non-parabolicity of the bands in both non-parabolic formulations. Particle flux and energy flux equations using a heated Fermi-Dirac distribution function are also presented, to allow for incorporation into existing device simulators.

Acknowledgements—This work was sponsored in part by ARPA through contract to NASA, NAGW-2753, the National Science Foundation through grant ECS-9313635 and by ARPA through the Phosphor Technology Center of Excellence, Contract MDA972-93-1-0030.

REFERENCES

1. G. Dresselhaus, *Phys. Rev.* **100**, 580 (1955).
2. R. H. Parmenter, *Phys. Rev.* **100**, 573 (1955).
3. E. O. Kane, *J. Phys. Chem. Solids* **1**, 249 (1957).
4. R. A. Stewart and J. N. Churchill, *Solid State Electron.* **33**(7) 819 (1990).
5. R. Thoma, A. Emunds, B. Meinerzhagen, H.-J. Peifer and W. L. Engl, *IEEE Trans Electron Device* **38**(6) 1343 (1991).
6. T. J. Bordelon, X.-L. Wang, C. M. Maziar and A. F. Tasch, *Solid State Electron.* **35**(2) 131 (1992).
7. D. Chen, E. C. Kan, U. Ravaioli, C.-W. Shu and R. W. Dutton, *IEEE Electron Device Letts.* **13**(1) 26 (1992).
8. E. M. Azoff, *J. Appl. Phys.* **64**(5) 2439 (1988).
9. D. L. Woolard, H. Tain, R. J. Trew, M. A. Littlejohn and K. W. Kim, *Phys. Rev. B* **44**(20) 11119 (1991).
10. D. Cassi and B. Riccò, *IEEE Trans. Electron Device* **37**(6) 1514 (1990).
11. N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Holt Rinehart and Winston, New York (1976).
12. J. Higman and K. Hess, *Solid State Electron.* **29**(9) 915 (1986).
13. A. W. Smith and A. Rohatgi, *IEEE Trans. Comput. Aided Des. Integrated Circuits* **12**(10) 1515 (1993).

COMPARISON OF NON-PARABOLIC HYDRODYNAMIC
SIMULATIONS FOR SEMICONDUCTOR DEVICES

A. W. SMITH and K. F. BRENNAN

School of Electrical and Computer Engineering and Microelectronics Research Center,
Georgia Institute of Technology, Atlanta, GA 30332-0250, U.S.A.

(Received 4 April 1995; in revised form 1 November 1995)

Abstract—Parabolic drift-diffusion simulators are common engineering level design tools for semiconductor devices. Hydrodynamic simulators, based on the parabolic band approximation, are becoming more prevalent as device dimensions shrink and energy transport effects begin to dominate device characteristics. However, band structure effects present in state-of-the-art devices necessitate relaxing the parabolic band approximation. This paper presents simulations of ballistic diodes, a benchmark device, of Si and GaAs using two different non-parabolic hydrodynamic formulations. The first formulation uses the Kane dispersion relationship in the derivation of the conservation equations. The second model uses a power law dispersion relation $\{(\hbar k)^2/2m = xW^\gamma\}$. Current-voltage relations show that for the ballistic diodes considered, the non-parabolic formulations predict less current than the parabolic case. Explanations of this will be provided by examination of velocity and energy profiles. At low bias, the simulations based on the Kane formulation predict greater current flow than the power law formulation. As the bias is increased this trend changes and the power law predicts greater current than the Kane formulation. It will be shown that the non-parabolicity and energy range of the hydrodynamic model based on the Kane dispersion relation are limited due to the binomial approximation which was utilized in the derivation. Copyright © 1996 Elsevier Science Ltd

NOTATION

\hbar	Planck's constant divided by 2π
k	reciprocal lattice vector
α	non-parabolicity factor under the Kane dispersion relation
m	carrier mass at the band edge, a constant
x, y	adjustable parameters for the power law dispersion relation
W	energy
T	electron temperature
\mathcal{F}_γ	Fermi integral of order γ divided by gamma function $(\gamma + 1)$
Γ	gamma function
μ_0	low field mobility value
μ	mobility value
E	electric field
ϵ_c	band edge parameters (potential, bandgap narrowing, affinity)
∇	gradient operator in physical space
v	electron velocity
S	electron energy flux
τ	relaxation times
p	hole concentration
n	electron concentration
K	Boltzmann's constant

INTRODUCTION

Drift-diffusion simulators and more recently hydrodynamic simulators are commonly being used as design tools for semiconductor devices. The hydrodynamic simulators are required as device dimensions

decrease and the energy transport effects begin to dominate device characteristics. These simulation tools are commonly based on the parabolic band approximation for the semiconductor materials. However, band structure effects present in state-of-the-art devices necessitate relaxing the parabolic band approximation. Several non-parabolic hydrodynamic models have been reported for homogeneous materials systems[1-5] using the Kane dispersion relationship[6]. Very limited device simulations based on these models have been presented[2,5], and fewer still have been compared to drift-diffusion simulations or parabolic hydrodynamic simulations of the same device[3]. The lack of simulation results may be due to the limited energy and non-parabolicity range over which the binomial Kane formulation is valid.

In a previous paper[7] we have presented two different non-parabolic hydrodynamic formulations suitable for device simulation; the parabolic form was also presented for reference. These forms resulted from the use of two different non-parabolic formulations for the dispersion relationship of the semiconductor in the derivation of the conservation equations. The first uses the classical non-parabolic Kane dispersion relationship, $\{(\hbar k)^2/2m = W(1 + \alpha W)\}$ [6], hereafter, also called the α formulation. The second uses a power law dispersion relation $\{(\hbar k)^2/2m = xW^\gamma\}$ [8]. The purpose of this paper is to present device simulations based on the non-parabolic hydrodynamic

equations presented in Ref. [7]. These simulations are compared to the parabolic hydrodynamic case for the same device, as well as the drift-diffusion simulations. Ballistic diodes, the benchmark device for hydrodynamic simulations, of silicon and gallium arsenide are simulated under a variety of bias conditions. Current-voltage characteristics of the three dispersion relations are compared utilizing the drift-diffusion approximation and the full hydrodynamic formulations.

This paper will proceed as follows: first, the ballistic diode structure is introduced, the doping and structure are the same for both material systems considered. The numerical aspects and physical models of the simulation code will then be presented. The material properties, including the non-parabolicity factors used, are presented for both Si and GaAs. Current-voltage characteristics for the silicon diode are discussed and explained using velocity and energy profiles. The discussion covers both drift-diffusion simulations and hydrodynamic simulations under the same conditions. Current-voltage characteristics for the GaAs diode are then presented. As in the case of the silicon simulations, velocity and energy profiles will be used to explain the results. In addition, mobility profiles will also be presented and explained in terms of the electric field and energy profiles. Energy profiles at high bias will be used to explain the breakdown in the α formulation.

BALLISTIC DIODE STRUCTURE

The ballistic diode structure examined is comprised of source and drain regions $0.2 \mu\text{m}$ in length doped at $5.0 \times 10^{17} \text{cm}^{-3}$ on both ends of a $0.4 \mu\text{m}$ thick $2.0 \times 10^{15} \text{cm}^{-3}$ n -type layer. This device structure is the benchmark for hydrodynamic (energy balance) simulations[9]. It was one of the first devices simulated using a full hydrodynamic simulator[10]. Although the device structure does not reflect all of the advanced physics included in the equations of Ref. [7] (Fermi-Dirac statistics, inhomogeneous materials), enough are present to highlight the applicability of the models. The bias is applied on the drain side, right hand side, of the device. The electrons enter the source and are swept out of the drain. The doping profiles and boundary conditions are assumed to be the same, regardless of the material system being simulated, silicon or GaAs.

NUMERICAL MODELS

The particle and energy continuity equations are the same as those derived by Woolard *et al.*[11]. The flux equations presented in this paper will be a subset of those presented in Ref. [7], in that all of the inhomogeneous material terms will be neglected. The particle flux equation (the derivation for all three cases is given in Ref. [7]) for the parabolic band case is

$$-n\bar{v} = \mu KT \left[\frac{\mathcal{F}_{1/2}}{\mathcal{F}_{-1/2}} \right] \nabla n + \mu n \nabla \epsilon_c + \frac{5}{2} \mu n KT \frac{\mathcal{F}_{3/2}}{\mathcal{F}_{1/2}} \left[1 - \frac{3}{5} \frac{\mathcal{F}_{1/2}^2}{\mathcal{F}_{-1/2} \mathcal{F}_{3/2}} \right] \frac{\nabla T}{T}. \quad (1)$$

The flux equation for the binomial α formulation is

$$-n\bar{v} = \mu KT \left[\frac{\mathcal{F}_{1/2}^2 + \frac{10}{4} \alpha KT \mathcal{F}_{3/2} \mathcal{F}_{1/2}}{\mathcal{F}_{1/2} \mathcal{F}_{-1/2} + \frac{15}{4} \alpha KT \left(\mathcal{F}_{1/2}^2 + \frac{\mathcal{F}_{-1/2} \mathcal{F}_{3/2}}{2} \right)} \right] \nabla n + \mu \left[\frac{\mathcal{F}_{1/2} - \frac{21}{4} \alpha KT \mathcal{F}_{3/2}}{\mathcal{F}_{1/2} + \frac{15}{4} \alpha KT \mathcal{F}_{3/2}} \right] n \nabla \epsilon_c + \frac{\mu n KT}{(\mathcal{F}_{1/2} + \frac{15}{4} \alpha KT \mathcal{F}_{3/2})(\mathcal{F}_{-1/2} + \frac{15}{4} \alpha KT \mathcal{F}_{1/2})} \times \left[\frac{5}{2} \mathcal{F}_{-1/2} \mathcal{F}_{3/2} - \mathcal{F}_{1/2}^2 - \frac{\mathcal{F}_{1/2}^3}{2(\mathcal{F}_{1/2} + \frac{15}{4} \alpha KT \mathcal{F}_{3/2})} \right] \frac{\nabla T}{T}. \quad (2)$$

The flux equation for the power law formulation is

$$-n\bar{v} = \frac{2\mu}{3xy^2} (KT)^{2-y} \frac{\left(\frac{y}{2} + 1\right) \Gamma\left(\frac{y}{2} + 1\right) \mathcal{F}_{y/2}}{\left(\frac{3}{2}y - 1\right) \Gamma\left(\frac{3}{2}y - 1\right) \mathcal{F}_{(3/2)y-2}} \nabla n + \frac{(2-y)\mu}{xy^2} (KT)^{1-y} \frac{\Gamma\left(\frac{y}{2} + 1\right) \mathcal{F}_{y/2}}{\Gamma\left(\frac{3}{2}y\right) \mathcal{F}_{(3/2)y-1}} n \nabla \epsilon_c + \frac{\mu n}{xy^2} (KT)^{2-y} \left[\frac{-y\left(\frac{y}{2} + 1\right) \Gamma\left(\frac{y}{2} + 1\right) \mathcal{F}_{y/2}}{\left(\frac{3}{2}y - 1\right) \Gamma\left(\frac{3}{2}y - 1\right) \mathcal{F}_{(3/2)y-2}} - \frac{(4+y) \Gamma\left(\frac{y}{2} + 2\right) \mathcal{F}_{(y/2)+1}}{3\Gamma\left(\frac{3}{2}y\right) \mathcal{F}_{(3/2)y-1}} \right] \frac{\nabla T}{T}. \quad (3)$$

The complete inhomogeneous energy flux equations for the three dispersion relations are listed as eqns (16)–(18) in Ref. [7]. The homogeneous material equations, which apply to the device structures of interest here, simplify to

$$-\bar{S} = \frac{5}{2} \mu n KT \left[KT \left[\frac{\mathcal{F}_{3/2}}{\mathcal{F}_{-1/2}} \right] \frac{\nabla n}{n} + \left[\frac{\mathcal{F}_{3/2}}{\mathcal{F}_{1/2}} \right] \nabla \epsilon_c + \frac{7}{2} KT \frac{\mathcal{F}_{5/2}}{\mathcal{F}_{1/2}} \left[1 - \frac{3}{7} \frac{\mathcal{F}_{1/2} \mathcal{F}_{3/2}}{\mathcal{F}_{-1/2} \mathcal{F}_{5/2}} \right] \frac{\nabla T}{T} \right] \quad (4)$$

$$\begin{aligned}
-S &= \frac{5\mu(KT)^2}{2} \\
&\times \left[\frac{\mathcal{F}_{3/2}(\mathcal{F}_{1/2} + \frac{15}{4}\alpha KT \mathcal{F}_{3/2}) - \frac{1}{4}\alpha KT \mathcal{F}_{1/2} \mathcal{F}_{5/2}}{(\mathcal{F}_{1/2} + \frac{15}{4}\alpha KT \mathcal{F}_{3/2})(\mathcal{F}_{-1/2} + \frac{15}{4}\alpha KT \mathcal{F}_{1/2})} \right] \nabla n \\
&+ \frac{5\mu n KT}{2} \left[\frac{\mathcal{F}_{3/2} - \frac{21}{4}\alpha KT \mathcal{F}_{5/2}}{(\mathcal{F}_{1/2} + \frac{15}{4}\alpha KT \mathcal{F}_{3/2})} \right] \nabla \epsilon_c \\
&+ \frac{5\mu n (KT)^2}{2(\mathcal{F}_{1/2} + \frac{15}{4}\alpha KT \mathcal{F}_{3/2})} \\
&\times \left[\frac{\frac{5}{2}\mathcal{F}_{5/2} - \frac{21}{8}\alpha KT \mathcal{F}_{7/2} + \frac{\mathcal{F}_{1/2}\mathcal{F}_{5/2} - \frac{3}{20}\alpha KT \mathcal{F}_{1/2}\mathcal{F}_{3/2}}{(\mathcal{F}_{1/2} + \frac{15}{4}\alpha KT \mathcal{F}_{3/2})}}{\frac{3\mathcal{F}_{3/2}(\mathcal{F}_{1/2} + \frac{21}{4}\alpha KT \mathcal{F}_{3/2}) - \frac{101}{8}\alpha KT \mathcal{F}_{1/2}\mathcal{F}_{5/2}}{2(\mathcal{F}_{-1/2} + \frac{15}{4}\alpha KT \mathcal{F}_{1/2})}} \right] \frac{\nabla T}{T}
\end{aligned} \quad (5)$$

$$\begin{aligned}
-S &= \frac{2\mu}{3xy^2} (KT_e)^{3-y} \frac{\left(\frac{y}{2} + 2\right) \Gamma\left(\frac{y}{2} + 2\right) \mathcal{F}_{(y/2)+1}}{\left(\frac{3}{2}y - 1\right) \Gamma\left(\frac{3}{2}y - 1\right) \mathcal{F}_{(3/2)y-2}} \nabla n \\
&+ \frac{(8-3y)\mu n}{3xy^2} (KT_e)^{2-y} \frac{\Gamma\left(\frac{y}{2} + 2\right) \mathcal{F}_{(y/2)+1}}{\Gamma\left(\frac{3}{2}y\right) \mathcal{F}_{(3/2)y-1}} \nabla \epsilon_c \\
&+ \frac{\mu n}{xy^2} (KT_e)^{3-y} \left[\frac{-y\left(\frac{y}{2} + 2\right) \Gamma\left(\frac{y}{2} + 2\right) \mathcal{F}_{(y/2)+1}}{\left(\frac{3}{2}y - 1\right) \Gamma\left(\frac{3}{2}y - 1\right) \mathcal{F}_{(3/2)y-2}} \right. \\
&\quad \left. + \frac{(6+y) \Gamma\left(\frac{y}{2} + 3\right) \mathcal{F}_{(y/2)+2}}{3\Gamma\left(\frac{3}{2}y\right) \mathcal{F}_{(3/2)y-1}} \right] \frac{\nabla T_e}{T_e}.
\end{aligned} \quad (6)$$

Equation (4) is the parabolic energy flux, while eqn (5) is for the alpha formulation and eqn (6) is for the power law formulation. The non-parabolic formulations have now been included in the two-dimensional device simulator STEBS-2D[12], which was originally a parabolic band hydrodynamic device simulator. The equations were discretized using the techniques in Refs [12 and 13]. For the exponential terms in the discretization equation, with factors composed of powers of the temperature, we have made the assumption that the position dependent temperature can be replaced by the average nodal temperature.

It has been shown by several researchers that the choice of the mobility model used in the hydrodynamic simulation can effect the shape of the velocity field relation[14] and the electronic thermal conductivity [15]. However, to provide a fair comparison with drift-diffusion simulations, which by their nature can not use an energy dependent mobility, we have chosen to use a field dependent mobility model for all of the hydrodynamic and drift-diffusion calculations presented here. The standard field dependent models for

Si[16] and GaAs[16] are used. The form used for GaAs captures the overshoot in the velocity-field relation. The value of the low field mobility, μ_0 depends upon the doping density and temperature of the device and is given in the next section. The equation used for μ_0 is given in Ref. [17] and will not be repeated here.

Another important physical model which is included in the simulation is a non-infinite heat sink for the lattice. This effect is included by using the energy exchange terms of Szeto[18] which are included as source/sink terms in the energy continuity equations (power density terms). The values for the relaxation times (τ_s) used in these expressions are listed in the next section.

The final numerical aspect of the model which must be addressed are the choices for the boundary conditions, both electrical and thermal in nature. Along both lateral sides of the ballistic diode symmetry boundary conditions are assumed for all of the variables. At the source and drain contacts, true ohmic conditions are used for the electrical contacts[16]. In the cases of the energy equations the temperatures (lattice and carriers) were set to 300 K at both ends of the device, implying full thermalization of the carriers.

MATERIAL PARAMETERS

The material parameters which are important for the successful simulation of the devices are the non-parabolicity factors and the energy relaxation times. For silicon using a bandgap of 1.124 eV and an effective mass of 0.326[19], the non-parabolicity factor can be estimated, from eqn (6) of Ref. [7], to be 0.4039 eV⁻¹. Fitting the band to the energy range (0 ≤ W ≤ 0.2) eV results in band parameters of $x = 1.15512$ and $y = 1.0439$ for the power law dispersion relation. The low field mobility values for Si are 1332.2 and 380.6 (cm² V⁻¹ s⁻¹) for doping densities of 2.0×10^{15} cm⁻³, and 5.0×10^{17} cm⁻³, respectively. The energy relaxation time in silicon is set to 0.2 ps for the electrons and 20 fs for the holes. Using GaAs with a bandgap of 1.424 eV and an effective mass of 0.070, the non-parabolicity factor is estimated to be 0.60736 eV⁻¹. Fitting the bands to the same range as in silicon, the power law non-parabolicity parameters for GaAs are $x = 1.2375$ and $y = 1.06489$. The low field electron and hole mobility values used for GaAs were 7940.9 and 2972 cm² V⁻¹ s⁻¹, respectively, for the doping densities of the ballistic diode. The energy relaxation time in GaAs is set to 0.1 ps for both electrons and holes.

SILICON BALLISTIC DIODE

Figure 1 displays the calculated current-voltage characteristics of the silicon ballistic diode for six different cases considered: three drift-diffusion simulations with different dispersion relations (parabolic,

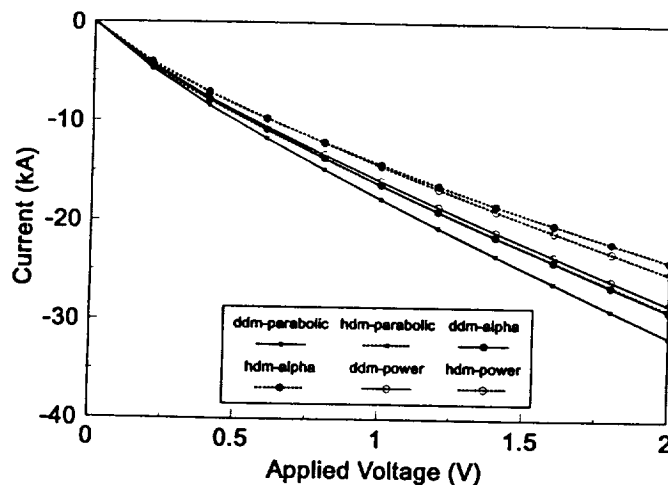


Fig. 1. Current-voltage relation of a silicon ballistic diode using the drift-diffusion approximation with three different energy dispersion relations and the full hydrodynamic formulation using the same three energy dispersion relations (ddm is drift-diffusion model, hdm is full hydrodynamic model).

α formulation, and power law) and three hydrodynamic simulations. The first observation is that the current calculated using the parabolic case is always greater than that determined using the non-parabolic models. This is true regardless of the use of the full hydrodynamic formulation, or just the drift-diffusion formulation. A cursory examination of eqns (1)–(3) shows that the non-parabolicity tends to reduce the effective mobility and diffusivity of the material system. The second observation is that the α formulation predicts greater current than the power law formulation at the same bias when using the drift-diffusion approximation. However, when the full hydrodynamic formulation is used the power law predicts greater current than the α formulation, at least at the higher values of applied bias. Figure 2 shows the calculated velocity profiles through the device for the

six cases considered. The figure confirms the current-voltage characteristics in Fig. 1, the non-parabolic velocities are all less than the parabolic case. Also, the velocity of the power law formulation under the drift-diffusion approximation is less than the α formulation. However, when the full hydrodynamic simulation is performed, the velocity of the power law form exceeds that obtained using the α formulation. Figure 3 displays the energy profile through the device, calculated using the temperature and the velocity from the current density. The small difference in the drift-diffusion energies is due to the differences in the velocities. These energy profiles are typical of those obtained for the ballistic diode [9,10]. However, notice that the non-parabolic formulations predict lower energy peaks than the parabolic formulation. This is contrary to that obtained in Ref. [3] for a similar structure. The cause

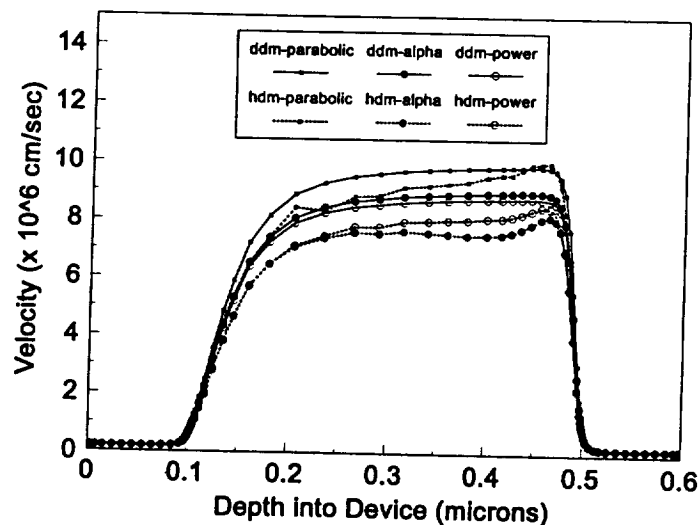


Fig. 2. Velocity profile of the silicon ballistic diode biased to 1 volt for the three energy dispersion relations under drift-diffusion conditions and hydrodynamic conditions.

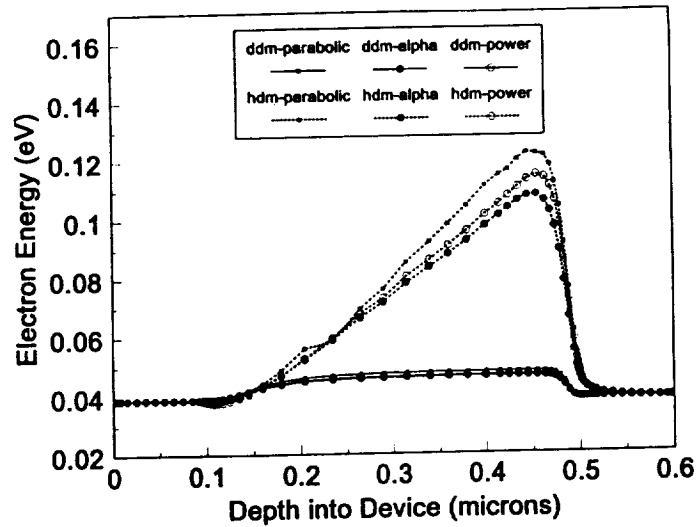


Fig. 3. Energy profile of the silicon ballistic diode biased to 1 volt for the three energy dispersion relations under drift-diffusion conditions and hydrodynamic conditions.

of this discrepancy is currently unknown. However, since the non-parabolicity factors tend to decrease the mobility and diffusivity, the energy trends displayed in Fig. 3 are consistent with all of the results presented here. These same trends in current, velocity and energy have also been observed when a self-consistent energy dependent mobility model[20] was used in the hydrodynamic simulations.

GALLIUM ARSENIDE BALLISTIC DIODE

Figure 4 displays the current-voltage characteristics simulated for the GaAs ballistic diode. Figure 4 shows the same trends as observed in Fig. 1 for silicon. The

non-parabolic formulations predict less current than the parabolic formulation using drift-diffusion or the full hydrodynamic form; the α formulation predicts more current than the power law form using the drift-diffusion approximation, but less current when the full hydrodynamic equations are simulated. However, at 0.2 volts applied bias the power law does predict slightly less current. To resolve this, the structure was simulated at finer voltage steps under low bias. The current-voltage characteristics from 0.0 to 0.4 volts for the two non-parabolic hydrodynamic formulations are shown as an insert to Fig. 4. The power law predicts smaller current values than the α formulation until the applied bias exceeds 0.32 volts, after which

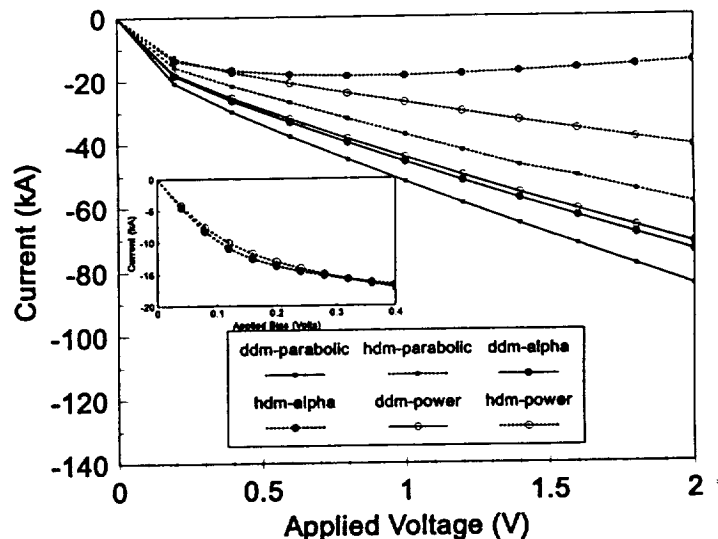


Fig. 4. Current-voltage relation of a GaAs ballistic diode using the drift-diffusion approximation with three different dispersion relations and the full hydrodynamic formulation using the same three dispersion relations. Insert shows current values of the full hydrodynamic formulation and the two non-parabolic dispersion relations at low bias.

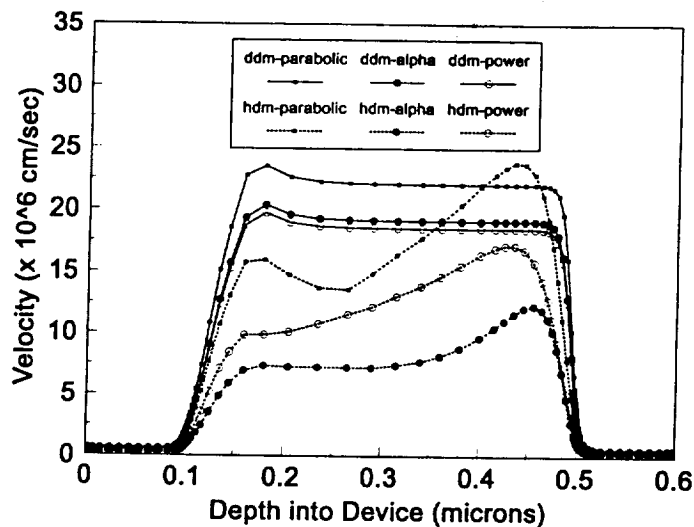


Fig. 5. Velocity profile of the GaAs ballistic diode biased to 1 volt for the three cases of dispersion relations under drift-diffusion conditions and hydrodynamic conditions.

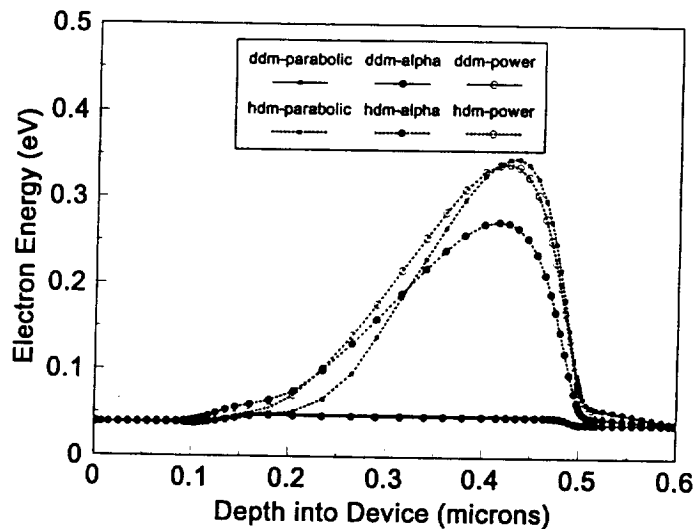


Fig. 6. Energy profile of the GaAs ballistic diode biased to 1 volt for the three cases of dispersion relations under drift-diffusion conditions and hydrodynamic conditions.

the power law predicts greater current. Another feature which is different than the case of silicon is that at higher biases the α formulation predicts lower current than at lower bias. It will be shown that this can be attributed to the breakdown of the binomial approximation used in the derivation.

The velocity and energy profiles for the GaAs structure under 1 volt bias are plotted in Figs 5 and 6. As for the case of the current, the same trend occurs as in silicon, the velocity predicted by the non-parabolic formulations is less than the parabolic case. The energy profile is different than that obtained for silicon, Fig. 3. First, near the source end of the device the non-parabolic forms predict greater energy than the parabolic case. At the drain end of the device the energy of the parabolic case does exceed the non-parabolic cases. However, the peak energy predicted

by the power law is quite close to the parabolic case, whereas the α form has a peak energy which is much lower.

Since the device is dominated by drift effects, it is beneficial to examine in detail the quantities which effect the drift component of the current. This includes the "effective" mobility and the field itself. Figure 7 is the "effective" mobility of the field term. This is the prefactor in front of the electric field term, nV_{ϵ_c} , in eqns (1)–(3). In the drift-diffusion simulations these factors are a constant, unless the field or doping changes the mobility. The field profiles for the hydrodynamic formulations differ only slightly from each other and the drift-diffusion cases. This is due to slight changes in carrier profiles caused by changes in the density of states and other parameters. The small differences in the field profile do not explain the large

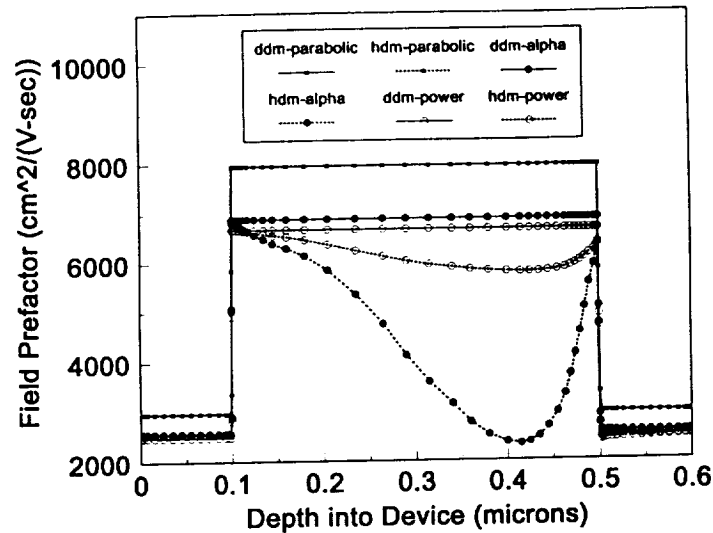


Fig. 7. "Effective" mobility on the field term in the flux equations. Under drift-diffusion conditions the mobility only changes due to doping and field effects. Under non-parabolic hydrodynamic conditions at high energy the "effective" mobility decreases.

variations in "effective" mobility observed in Fig. 7. The doping profile does play a role in changing the mobility as displayed in Fig. 7 for the drift-diffusion cases where there is an abrupt change in mobility when the doping changes. Notice that the "effective" mobility of the power law form is lower than the α form in the drift-diffusion model, as shown in Fig. 7. This leads to the lower current values under the drift-diffusion approximation for the power law. In contrast, under full hydrodynamic simulation the α formulation's "effective" mobility is lower, leading to smaller current values. The current-voltage trend observed in the insert of Fig. 4 can now be explained using Fig. 7. At low bias the energy of the carriers has not risen significantly above the drift-diffusion case,

the "effective" mobility of the α formulation is greater than the power law formulation and for the same field this leads to higher current using the α formulation. As the bias is increased substantial heating of the carriers occurs, the "effective" mobility under the α formulation decreases more than the power law form leading to lower current values.

As previously stated, the current simulated under the hydrodynamic α formulation decreases as the bias is increased. This can be attributed to the breakdown in the binomial expansion used in the derivation of the closed form coefficients. Azoff[21] predicted that the binomial expansion on the diffusion term becomes inaccurate when the condition $2\alpha W < 1$ no longer holds. In Azoff's work[21] non-parabolicity effects

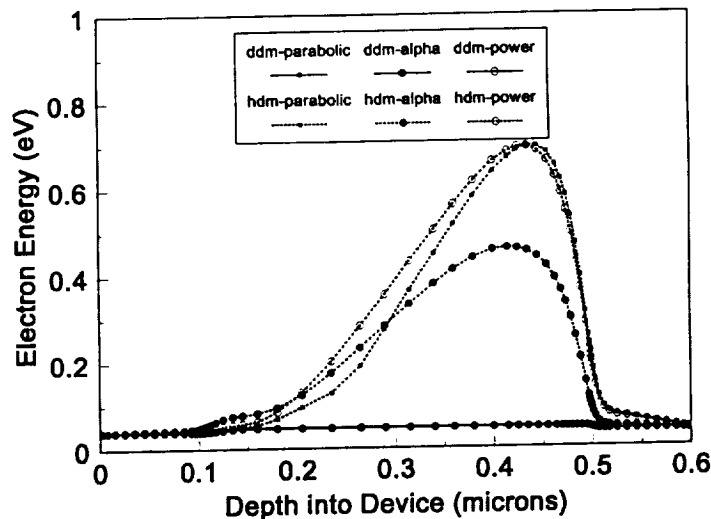


Fig. 8. Energy profile of the GaAs ballistic diode biased to 2 volts for the three cases of dispersion relations under drift-diffusion conditions and hydrodynamic conditions. The validity conditions of the binomial expansion in the α formulation have been exceeded.

case,
much

, it is
which
ludes
ure 7
s the
, in
these
ping
dro-
each
e to
es in
mall
arge

on the field were not taken into account. For Azoff's value of $\alpha = 0.5823$ this condition implies a hot carrier energy limit of 0.859 eV, i.e. the average energy of the carriers should not exceed 0.859 eV. In this paper, we have used an $\alpha = 0.60736$ which implies a hot carrier limit of 0.823 eV for the diffusion term. However, in the derivation of the field dependent term within the present work, where non-parabolicity is included [Ref. [7], Equation (13)], the use of the binomial expansion in simplifying the denominator leads to an extra factor of two. In this case, the binomial expansion becomes inaccurate yielding unreliable results when the condition $4\alpha W < 1$ is invalid. For the choice of $\alpha = 0.60736$, this restriction implies a hot carrier energy limit of 0.4115 eV. The peak average energy in Fig. 6 is less than 0.4115 eV for all three cases of the dispersion relation and the difference in peak energies for all three formulations is not large. However, at a two volt bias the average energy of the parabolic and power law formulations exceeds the limiting value reaching values of approximately 0.7 eV, as shown in Fig. 8. On the other hand the peak average energy of the α formulation fails to track the increase in the energy predicted by the parabolic and power law models. As can be seen from Fig. 8, the α formulation predicts a much lower peak average energy of 0.5 eV, still above the range of energies for which this model is suspect. It is clear then, that at higher applied voltages and consequently high average energies, use of the α formulation containing a binomial expansion leads to inaccurate predictions of the key macroscopic observables, i.e. average energy and current.

CONCLUSIONS

In this paper, ballistic diode simulations using two different non-parabolic formulations of the energy dispersion relation within a hydrodynamic model have been performed. These calculations have been compared to those made with non-parabolic drift-diffusion models and a parabolic hydrodynamic model. The first non-parabolic formulation of the energy utilizes the Kane dispersion relation. In order to produce closed form coefficients in the transport equations, thereby making the calculation numerically tractable, it is essential that the terms be simplified using a binomial expansion. Unfortunately, the resulting binomial expansion is only valid for a narrow, low energy range at modest values of non-parabolicity. As a result, the non-parabolic formulation based on the Kane dispersion relation is only suitable for calculation within a narrow energy range.

In contrast, the power law formulation for the energy dispersion relation has a much greater range of energy over which it provides a valid description of the transport coefficients. This can be attributed to the observation that the transport coefficients at high energy become non-physical when the binomial expansion, Kane dispersion relation, α formulation is

employed, while in the power law formulation no non-physical effects are observed.

For devices and material systems considered here, it is found that the use of the non-parabolic as opposed to the parabolic dispersion relation always leads to a lower calculated current in both the drift-diffusion and hydrodynamic simulations. The calculated carrier energy and velocity are similarly lower when the non-parabolic model is employed. These variables are lower because the non-parabolicity reduces the effective mobility and diffusivity of the material, thereby lowering the calculated current.

The relationship between the two non-parabolic models, the power law and the α formulation, is somewhat more complicated. The power law formulation predicts smaller current values than the α formulation under the drift-diffusion approximation in both silicon and GaAs. However, when the full hydrodynamic formulation is used, the power law predicts greater current than the α formulation, at least at high bias. At low bias, the power law predicts a slightly lower current than the α formulation. These results can be attributed to the behavior of the "effective" mobility, since in the ballistic diodes considered here the current is dominated by drift effects. When the drift-diffusion model is employed, the "effective" mobility (defined as the prefactor in front of the electric field term) of the power law form is lower than that of the α form, leading to a lower calculated current. In contrast, when the full hydrodynamic model is employed, the "effective" mobility of the power law form is greater than that of the α form, resulting in a larger calculated current. The observation of the power law initially yielding a lower current at low bias, but yielding a higher current at high bias as compared to the α form, can be explained as follows. At low bias, the devices are close to equilibrium and the carriers are relatively cold. Consequently, the system responds as in the drift-diffusion case, resulting in a greater current for the α case than the power law case. At higher biases, the effects of carrier heating are more important, and the situation reverses, since the full hydrodynamic results dominate the "effective" mobility.

Acknowledgements—This work was sponsored in part by ARPA through contract to NASA, NAGW-2753, the National Science Foundation through grant ECS-9313635, and by ARPA through the Phosphor Technology Center of Excellence, Contract MDA972-93-1-0030.

REFERENCES

1. R. A. Stewart and J. N. Churchill, *Solid St. Electron.* **33**, 819 (1990).
2. R. Thoma, A. Edmunds, B. Meinerzhagen, H.-J. Peifer and W. L. Engl, *IEEE Trans. Electron. Devices* **38**, 1343 (1991).
3. T. J. Bordelon, X.-L. Wang, C. M. Maziar and A. F. Tasch, *Proc. IEDM* pp. 353 (1990).
4. T. J. Bordelon, X.-L. Wang, C. M. Maziar and A. F. Tasch, *Solid St. Electron.* **35**, 131 (1992).

5. D. Chen, E. C. Kan, U. Ravaioli, C.-W. Shu and R. W. Dutton, *IEEE Electron. Device Lett.* **13**, 26 (1992).
6. E. O. Kane, *J. Phys. Chem. Solids* **1**, 249 (1957).
7. A. W. Smith and K. F. Brennan, *Solid St. Electron.* (submitted).
8. D. Cassi and B. Riccò, *IEEE Trans. Electron. Devices* **37**, 1514 (1990).
9. S. Ramaswamy and T.-W. Tang, *IEEE Trans. Electron. Devices* **41**, 76 (1994).
10. R. K. Cook and J. Frey, *IEEE Trans. Electron. Devices* **28**, 951 (1981).
11. D. L. Woolard, H. Tain, R. J. Trew, M. A. Littlejohn and K. W. Kim, *Phys. Rev. B* **44**, 11119 (1991).
12. A. W. Smith, Ph.D. dissertation, Georgia Institute of Technology, May (1992).
13. A. W. Smith and A. Rohatgi, *IEEE Trans. Comput. Aid. Des. Integrated Circuits* **12**, 1515 (1993).
14. A. D. Sandomnikov and D. J. Roulston, *COMPEL* **12**, 245 (1993).
15. K. Souissi, F. Odeh, H. H. K. Tang and A. Gnudi, *COMPEL* **13**, 439 (1994).
16. S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Springer, New York. (1984).
17. P. A. Basore, PC-1D installation manual and user's guide version 3.0, Sandia National Laboratories, Albuquerque, NM (1991).
18. S. Szeto and R. Reif, *Solid St. Electron.* **32**, 307 (1989).
19. M. A. Green, *J. Appl. Phys.* **67**, 2944 (1990).
20. B. Pejčinović, T.-W. Tang, S.-C. Lee and D. H. Navon, *IEEE Trans. Electron. Devices* **39**, 2021 (1992).
21. E. M. Azoff, *J. Appl. Phys.* **64**, 2439 (1988).

ation no
red here,
opposed
ads to a
diffusion
d carrier
hen the
variables
uces the
naterial,

arabolic
is some-
ulation
ulation
n silicon
ynamic
greater
gh bias.
y lower
can be
obility,
current
ffusion
defined
orm) of
form,
ntrast,
ed, the
greater
ulated
nitially
ding a
form,
levices
atively
in the
nt for
iases,
t, and
namic

part
3, the
13635,
ter of

n. 33,

Peifer
1343

A. F.

A. F.

SECTION V:

ACOUSTIC CHARGE TRANSPORT DEVICE

5.0 Acoustic Charge Transport Device

We present the figure of the ACT imager architecture here again to clarify how we segmented the problem. From the substrate architecture shown it should be noted that the photogenerated charge is transferred in the GaAs quantum well sandwiched between two AlGaAs layers. Control of mobile carriers is critical to this operation and so we developed a layer structure which would be best suited for ACT. An earlier heterostructure ACT approach was pioneered by a group at United Technologies with the principal advantage that no biasing was required. It turns out, however, that in their initial structure the depletion of the charge transport layer depended on the surface state concentration being within a particular range of values. We developed a layer structure for which the depletion of the charge transport layer would be relatively impervious to the surface state concentration. This will be discussed in detail in Section 5.1. One thing that is also something of a problem for ACT devices is the high amount of RF power that is typically required for charge transport. This is because GaAs is weakly piezoelectric with a value for the piezoelectric coupling, K^2 , of 0.07% which is less than that of Quartz. This high RF power causes a number of other problems such as electromigration and acoustomigration in the SAW interdigital transducer (IDT) metallization. In addition, the need for high acoustic power in the transport channel results in an increased amount of chip real estate taken up by the IDT and the reflector behind it. There are several ways in which we sought to deal with these problems. One was to investigate the use of a thin film piezoelectric layer of Zinc Oxide (ZnO) to be deposited on top of the GaAs ACT device. Due to the much higher value of K^2 for ZnO, this dramatically decreases the amount of RF power driving the SAW IDT. This work is described in Section 5.2. Another approach we took to solving this problem was to investigate SAW acoustoelectric convolvers. With this approach there would be

no need for an IDT and the charge transport would be powered by a modest DC Voltage (~2 Volts) applied to the acoustelectric amplifier and current in the mA range. This work is described in Section 5.3.

It is perhaps at first not so evident but the photodetector array presents an acoustic waveguide structure which the SAW will have to propagate through. In order to better understand this problem and provide a design methodology we developed the stack matrix theory approach for the design of SAW waveguides and this work is described in Section 5.4. This technique has been adopted as the industry standard for the design of SAW Waveguide coupled resonator filters now used for IF sections of wireless systems.

5.1 p-HACT Device Architecture

In this section we describe the ACT device architecture which we developed to overcome some of the manufacturing difficulties associated with the heterostructure ACT architecture. For this section of the report we have attached the paper which most extensively describes the work:

Smith, A.W., Kenney, J.S., Hunt, W.D., Brennan, K.F., Benz, R. and Summers, C.J., "Theoretical Calculations of Charge Confinement in a pn⁺np Heterojunction Acoustic Charge Transport Device," *IEEE Transactions on Electron Devices*, vol. 42, no. 5, pp. 977-990, May 1995.

The problem that this work addresses can best be understood by looking at Figure 3 of the attached paper. From this one can see that the Conduction Band edge varies considerably for HACT devices for only modest variations in the surface state concentration. It is well known that the surface state concentration cannot be held to a specific value and that a more effective architecture is one for which the devices function well for a surface state density held below a specific value. The problem with the HACT architecture shown in Figure 2 of the paper is that it depends on a specific value of surface state concentration being attained. We developed the architecture in Figure 4 to be a much better alternative to immunity from surface state concentration variations.

Theoretical Calculations of Charge Confinement in a $pn^{-}np$ Heterojunction Acoustic Charge Transport Device

Arlynn W. Smith, J. Stevenson Kenney, William D. Hunt, Kevin F. Brennan,
Senior Member, IEEE, Rudy Benz, and Christopher J. Summers, Member, IEEE

Abstract— An alternative structure for heterojunction acoustic charge transport (HACT) devices has been devised and analyzed. The structure uses a $pn^{-}np$ doping profile near the surface of the device to create a charge transport layer and provide top vertical confinement. This is contrary to previous n-type HACT structures which rely on residual surface states and a heterojunction discontinuity for the same functions. The use of the $pn^{-}np$ doping as the channel depletion mechanism makes the device insensitive to the residual surface state density, thus providing a more robust design. In addition, the use of the back np junction enables widening of the transport layer thereby increasing the amount of charge that can be transported by the acoustic wave. As a result of the increased charge capacity it is expected that the $pn^{-}np$ ACT device will exhibit a greater dynamic range and current than previous HACT designs. The analysis of the device structure is accomplished herein using a two dimensional hydrodynamic simulation code, Semiconductor Total Energy Balance Simulator in two Dimensions (STEBs-2D), which has been modified to account for the potential created by the surface acoustic wave. The calculated results indicate that an order of magnitude enhancement in charge capacity is possible using the new structure. Transfer efficiency calculations for several different lifetimes in the transport layer show high efficiency values, greater than 79% efficient with a Shockley Read Hall lifetime of 10 nsec.

NOMENCLATURE

ϵ_r	Relative permittivity of semiconductor.
ϵ_0	Permittivity of free space.
q	Electronic charge.
ψ	Electrostatic potential.
p	Hole concentration.
n	Electron concentration.
N_d^+	Ionized donor concentration.
N_a^-	Ionized acceptor concentration.
Γ_{SAW}	Effective doping due to surface acoustic wave.
t	Time.
$v_{n,p}$	Velocity of electrons and holes.
R, G	Recombination and Generation.

$\mu_{n,p}$	Electron and hole mobility.
k_b	Boltzmann constant.
$T_{e,h}$	Electron or hole temperatures.
ϵ_c	Conduction band edge (includes bandgap narrowing).
ϵ_v	Valance band edge (includes bandgap narrowing).
χ	Electron affinity.
E_g	Semiconductor band gap.
$m_{e,h}^*$	Electron and hole effective mass.
Λ_c	Ratio of Fermi integrals $\{F_{1/2}(\eta_c)/F_{-1/2}(\eta_c)\}$.
η_c	Reduced energy of the conduction band.
Λ_v	Ratio of Fermi integrals $\{F_{1/2}(\eta_v)/F_{-1/2}(\eta_v)\}$.
η_v	Reduced energy of the valance band.

I. INTRODUCTION

HETEROJUNCTION acoustic charge transport (HACT) devices are promising candidates for analog signal processors and rf memories. The HACT device structure is just one of the two major forms of acoustic charge transport (ACT) devices, bulk and surface types. In bulk ACT structures charge is confined in the vertical direction (~ 5 to $6 \mu\text{m}$ deep) using a bias potential created between a reverse-biased Schottky barrier and an applied back-gating potential. In HACT devices vertical confinement ($< 1 \mu\text{m}$ deep) is provided by discontinuities in the conduction band edge due to material changes. In both HACT and ACT devices lateral confinement and charge transport is accomplished using the potential created by a surface acoustic wave generated within the semiconductor. A schematic diagram of acoustic charge transport in a representative HACT structure is shown in Fig. 1. As can be seen from Fig. 1, the surface acoustic wave generated by a transducer on the surface of the semiconductor propagates through the structure pushing charge from the input contact over some distance for later extraction at an output contact.

The HACT structure offers several important advantages over a conventional buried channel ACT. The HACT device has less power dissipation, utilizes simpler channel isolation techniques, is self-biasing, and enables possible direct integration with other planar technologies [1]. These advantages are realized due to the shallower depth of confinement for the charge packet. Less input power is required because more efficient use of the surface acoustic wave (SAW) potential

Manuscript received May 2, 1994; revised December 21, 1994.

A. W. Smith, J. S. Kenney, W. D. Hunt, and K. F. Brennan are with the School of Electrical and Computer Engineering and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

R. Benz and C. J. Summers are with the Quantum Microstructures Branch, Georgia Tech Research Institute, Georgia Institute of Technology, Atlanta, GA 30332-0800 USA.

IEEE Log Number 9410170.

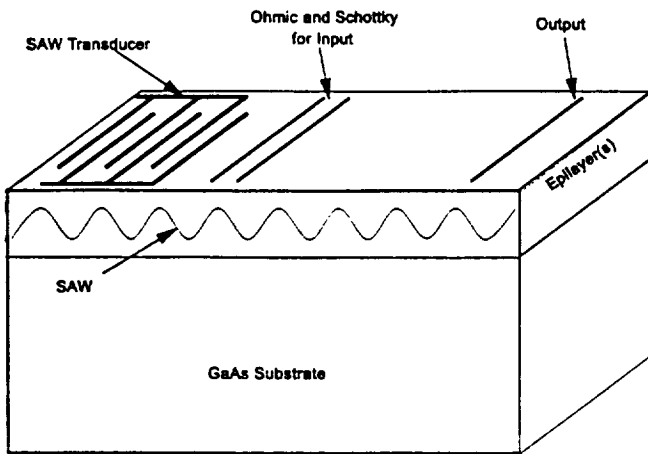


Fig. 1. Schematic diagram of a surface acoustic wave propagating in an epilayer structure.

is obtained from the shallower charge confinement. Channel isolation can be provided by mesa etching or proton isolation which would not be possible at a depth of $5\ \mu\text{m}$. However, these advantages do not come without some penalty. In the conventional HACT (n-HACT) structure, shown in Fig. 2, the self-biasing of the device relies on residual surface states and very critical control of the doping in the layer structure. The one dimensional conduction band diagram, without any acoustic potential, for three different cases of residual surface charge for the n-HACT structure is displayed in Fig. 3, the surface charge densities are $-1.0 \times 10^{12}\ \text{cm}^{-2}$, $-1.5 \times 10^{12}\ \text{cm}^{-2}$, and $-5.0 \times 10^{12}\ \text{cm}^{-2}$. The fermi level is at zero volts. Notice that at the lowest residual surface concentrations the channel is modulation doped by the dopants in the top AlGaAs layer. At moderate values of surface charge [2] the channel is depleted as required. This is the first surface state density for which we obtained charge confinement. As the surface charge increases the channel becomes more depleted. The depth of the depletion region is controlled by the density of the surface states and the doping concentration in the layers below. The design goal is to have the doping concentration in the AlGaAs layer slightly larger than the fixed surface charge [1]. Once the surface states have been satisfied this will then provide a small amount of mobile charge for transport [1]. A major problem is that the resulting value of surface charge is not known *a priori* i.e., different processing conditions or even different atmospheric conditions could change the final residual surface concentration.

Small variations in the doping concentrations within the layers can also upset the delicate charge balancing with the surface state density. Any of these deviations from perfect balance can cause large variations in the device performance. Relying on surface charge to deplete the channel layer also introduces a limit on the width of the transport layer, the surface charge acts over a small distance unless the material is lightly doped. A smaller transport layer thickness limits the charge capacity of the acoustic wave packet thereby decreasing the dynamic range of the device. Therefore, a need exists for a HACT device design which is insensitive to the residual surface states and has performance characteristics comparable

20 nm	GaAs	N-type NID
70 nm	AlGaAs with 30% Aluminum	N-type 2.0×10^{17}
40 nm	GaAs	N-type NID
100 nm	AlGaAs with 30% Aluminum	N-type NID
GaAs Substrate		N-type NID

Fig. 2. Epitaxial layer structure of a conventional n-type heterojunction acoustic charge transport device (n-HACT).

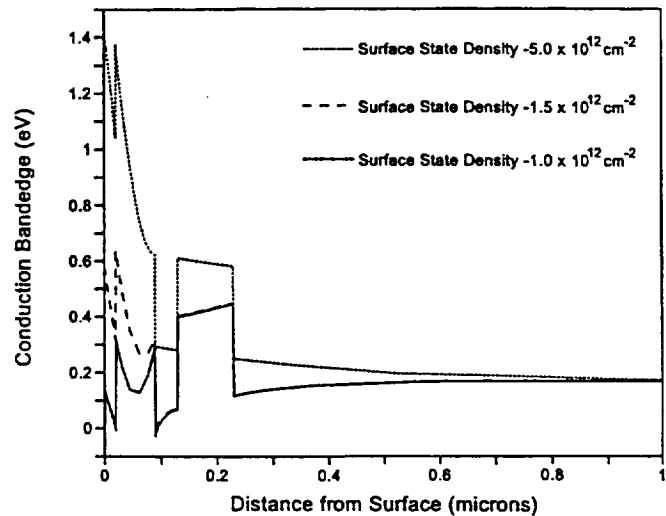


Fig. 3. One dimensional conduction band diagrams of the n-type HACT device with three different residual surface state concentrations, no acoustic potential.

to or greater than the current HACT structures which rely on the surface states and have limited channel widths. A structure which utilizes a pn junction at the surface to isolate the channel from surface states was proposed in reference [3]. However, the structure was never analyzed to determine the necessary doping profiles or the device characteristics. The structure proposed in this paper is an extension of the pn HACT configuration first described in reference [3].

The purpose of this paper is to introduce and analyze an alternative HACT structure, hereafter called p-HACT, which is relatively insensitive to changes in residual surface concentration and to fluctuations in the doping concentration of the epitaxial layers. The proposed device structure utilizes a pn junction at the front surface to provide isolation from the surface charge and to create the top vertical confinement. A lightly doped n-type channel layer is then separated from the substrate by a p-type AlGaAs layer. The p-type AlGaAs serves two purposes: First, the band edge discontinuity provides the bottom part of the vertical confinement. Second, the p-type doping serves to complete the depletion of the transport layer. By using dopants to deplete the transport layer, channel thicknesses an order of magnitude larger than those obtained

50 nm		P-type 2.15×10^{17}
50 nm		N-type 1.85×10^{17}
300 nm	GaAs	N-type 1.0×10^{15}
100 nm	AlGaAs with 30% Aluminum	P-type 1.0×10^{18}
GaAs Substrate		N-type 5.0×10^{15}

Fig. 4. Epitaxial layer structure of the alternative p-type heterojunction acoustic charge transport device (p-HACT).

in current HACT devices can be achieved, sensitivity to the surface states is reduced, and processing variations can be avoided.

The remainder of the paper will discuss the two dimensional device simulation code, Semiconductor *Total Energy Balance Simulator* in two Dimensions (STEBs-2D), used for modeling the performance of the HACT structures. The charge packet size and shape for the two different HACT designs at one specific acoustic power level will then be compared for three different residual surface charge concentrations. Finally, transient simulations using several values of carrier lifetime in the channel layer provide a measure of the maximum charge transfer efficiency of the proposed structure.

II. DEVICE DESIGN

The layer structure and doping concentrations in the new p-type HACT, hereafter referred to as p-HACT, is shown in Fig. 4. The design consists of a top layer of p-type GaAs doped to $2.15 \times 10^{17} \text{ cm}^{-3}$ with a thickness of 50 nm, 50 nm n-type GaAs doped to $1.85 \times 10^{17} \text{ cm}^{-3}$ follows the cap layer. The channel layer is 300 nm of n-type GaAs doped to $1.0 \times 10^{15} \text{ cm}^{-3}$. The bottom vertical confinement layer is 100 nm of p-type $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ with a doping concentration of $1.0 \times 10^{16} \text{ cm}^{-3}$. The substrate is n-type GaAs doped at $5.0 \times 10^{15} \text{ cm}^{-3}$. Alternatively a $2 \mu\text{m}$ GaAs buffer layer of the same doping concentration on a semi-insulating substrate can be used. The one dimensional conduction band diagram of the p-HACT, without the superimposed acoustic potential, for three different residual surface state concentrations is shown in Fig. 5, the surface state densities are $-5.0 \times 10^{11} \text{ cm}^{-2}$, $-1.0 \times 10^{12} \text{ cm}^{-2}$, and $-5.0 \times 10^{12} \text{ cm}^{-2}$ which differ from those used in the n-HACT simulation. Comparing Figs. 3 and 5 it is immediately apparent that the p-HACT structure is much less sensitive to fluctuations in the top surface charge; the band diagram does not change as the surface charge varies. Also note that the width of the transport layer has been increased by almost an order of magnitude, from 40 nm to 300 nm.

The maximum doping concentration in the two HACT structures is approximately the same for several reasons. First, it was found that when the doping concentrations in the first two layers was reduced the p-HACT became more sensitive to the surface state density. Similar behavior occurs for the n-HACT structure. Since the design goal is to produce a device insensitive to the surface state concentration, the doping

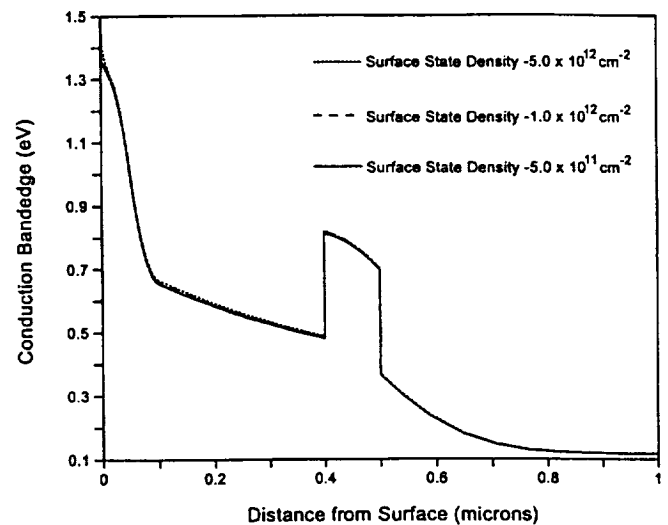


Fig. 5. One dimensional conduction band diagrams of the p-HACT device with three different residual surface state concentrations, no acoustic potential.

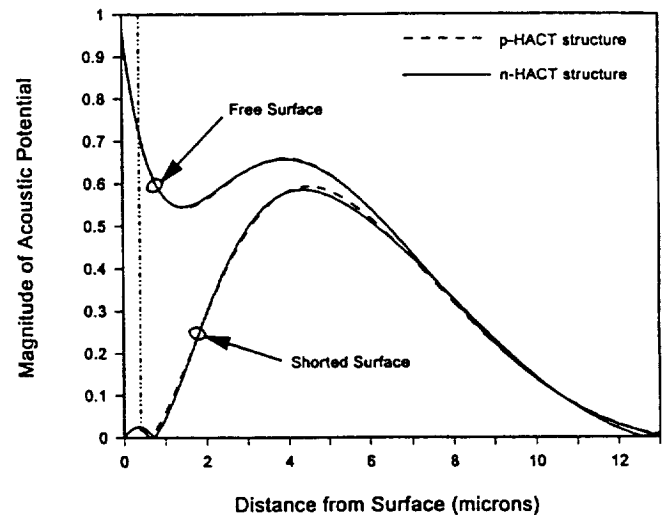


Fig. 6. Magnitude of the acoustic potential in the n-HACT and p-HACT structures, the transport channels are located between the origin and the vertical line. Two different boundary conditions are displayed: a free surface boundary condition and a shorted (metallized) boundary condition.

concentrations in the top layers must remain fairly high. However, the doping concentration in these layers can not be increased to arbitrarily high values in either structure owing to the fact that the top surface of the HACT structure can not be highly conductive [1]. A highly conductive surface, such as a metal or highly doped semiconductor, will short out the surface acoustic wave propagating in the underlying semiconductor. This will be discussed at length in the next section. In the case of the n-HACT structure the doping concentration in the AlGaAs cap layer can not be raised too high because the layer will begin modulation doping the transport layer, similar to a HEMT. As the number of carriers increases in the channel due to the modulation doping the acoustic wave will be damped out, the perturbation due to the displacement charge from the surface acoustic wave will become too small compared to the background carrier concentration. In the p-HACT device,

$$p\bar{\nu}_p = \mu_h \left[-k_b T_h \Lambda_v \nabla p - p \nabla (\epsilon_v - \chi - E_g) + \frac{3}{2} p k_b T_h \Lambda_v \frac{\nabla m_h^*}{m_h^*} \right]. \quad (5)$$

high doping creates a very strongly depleted channel but the perturbation due to the acoustic wave is not great enough to cause significant lateral confinement. For this case the heavy doping strongly pins the energy bands in the channel region.

III. MODEL DESCRIPTION

The ability to model the HACT structure depends upon the assumption that the acoustic properties of the material can be treated as a perturbation to the electrical properties under specific conditions. The SAW potential is converted to an effective doping concentration and superimposed on the background concentration of the structure. This is more realistic than superimposing the SAW potential directly on the device potential because at the surface and deep within the substrate the SAW contributes little to the overall potential.

A. Electrical Model

The fundamental basis of our model is a two dimensional hydrodynamic device simulation code, STEBS-2D [4]. For the simulations presented here, the lattice temperature effects and electronic heat transfer have been neglected. The resulting partial differential equations describing the system and solved within the context of the hydrodynamic simulator are

$$\nabla(\epsilon_r \nabla \psi) = \frac{q}{\epsilon_o} (p - n + N_d^+ - N_a^- + \Gamma_{SAW}) \quad (1)$$

$$\left(\frac{\partial n}{\partial t} \right) + \nabla \cdot (n \bar{\nu}_n) = -R + G \quad (2)$$

$$\left(\frac{\partial p}{\partial t} \right) + \nabla \cdot (p \bar{\nu}_p) = -R + G \quad (3)$$

$$n \bar{\nu}_n = \mu_n \left[-k_b T_e \Lambda_c \nabla n + n \nabla (\epsilon_c - \chi) + \frac{3}{2} n k_b T_e \Lambda_c \frac{\nabla m_c^*}{m_c^*} \right] \quad (4)$$

(see (5) shown at the top of the page.)

Equation (1) is Poisson's equation for the potential and includes the effective doping from the surface acoustic wave, Γ_{SAW} , as will be described. Equations (2) and (3) are the electron and hole current continuity equations. The electron and hole flux equations are given in (4) and (5). On the right hand side of (2) and (3), the generation/recombination terms (R and G) account for band to band, Auger, and Shockley-Read-Hall (SRH) recombination. The partial differential equations are discretized using the control volume method as described in [4], [5]. All material parameters i.e. bandgap, mobility, density of states, recombination parameters are determined using the formulas stated in PC-1D [6]. In addition, a thermionic-field emission boundary condition [7] for the currents is implemented at each material discontinuity. Interface states

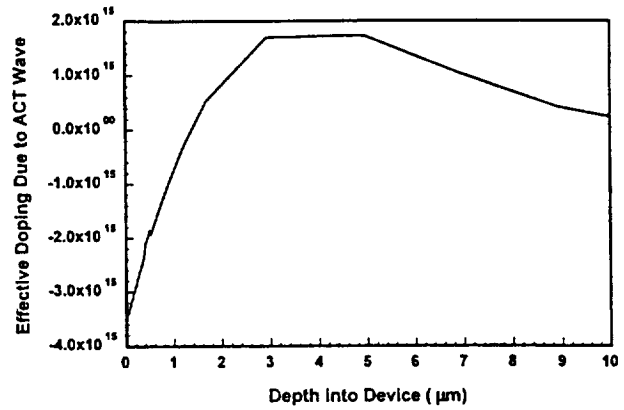


Fig. 7. Maximum value of the displacement charge created by the acoustic wave as a function of depth into the p-HACT structure. The discontinuities are caused by abrupt changes in material permittivity.

and recombination centers were assumed to be negligible at the material discontinuities. Equations (1)–(5), along with the appropriate boundary conditions for the device domain, are solved for the three fundamental variables ψ , n , and p .

In the simulations presented here four different boundary conditions are needed. The first is a grounded ohmic contact [8] at the back surface of the device. The second type of boundary condition is the symmetry boundary condition [8]. This is applied on both lateral sides of the simulation domain. This choice of boundary conditions will be explained further in the next section on modeling the acoustic wave. On the top surface two different types of boundary conditions are used. In the one dimensional simulations without the surface acoustic wave, Figs. 3 and 5, a free surface boundary condition is employed for the electrostatic potential. Its mathematical form is

$$\epsilon_{semi} \frac{\partial \psi}{\partial X_i} = Q_{int} \quad (6)$$

The parameter ϵ_{semi} is the semiconductor permittivity and Q_{int} is the surface charge density and varies in the range -5×10^{11} to $-5 \times 10^{12} \text{ cm}^{-2}$. This range of surface charge density corresponds to approximately 1% of the dangling bonds being charged. Even though the SAW is expected to alter the surface state occupancy (Q_{int}), this has been neglected in the model due to the lack of information concerning the charged state profile. The boundary conditions for the carrier concentrations depend upon satisfying the continuity equations at the surface with respect to the surface recombination rate as,

$$J_n = -q R_{surf} = -q \frac{S_n S_p (np - n_i^2)}{S_n(n+n_1) + S_p(p+p_1)} \cdot \delta(x) \quad (7)$$

$$J_p = q R_{surf} = q \frac{S_n S_p (np - n_i^2)}{S_n(n+n_1) + S_p(p+p_1)} \cdot \delta(x)$$

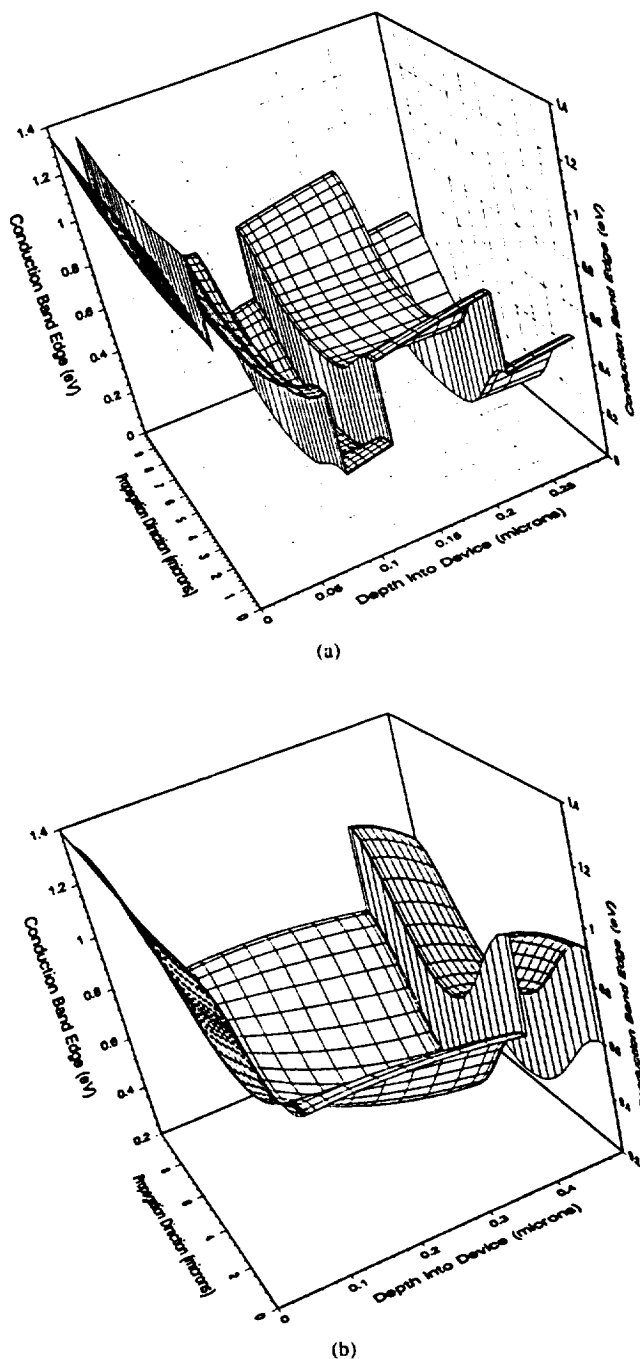


Fig. 8. Two dimensional conduction band diagrams for the (a) n-HACT structure, and (b) p-HACT structure with an applied surface acoustic wave. Only the epitaxial layer are displayed in each case.

In the one dimensional simulations the value of the surface recombination rates (R_{surf}) used in (7) has been set to zero. Setting R_{surf} to zero is justified because the surface is not driven out of equilibrium. Therefore $np = n_i^2$ and no matter what the actual surface recombination velocities are the surface recombination rate evaluates to zero [8]. These one dimensional simulations then provided the value of the fermi level with respect to the intrinsic level which was used in the two dimensional simulations.

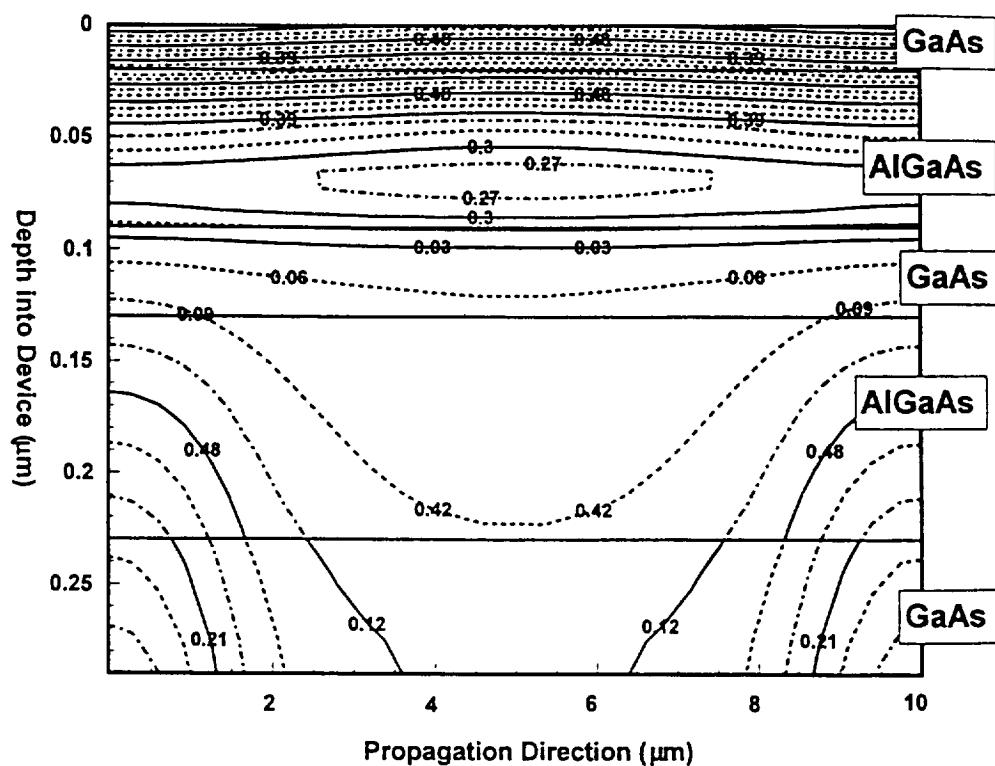
In the two dimensional simulations with the SAW the free surface boundary conditions given by (6) and (7) in the

one dimensional case could be used with the modification that an extra term accounting for the acoustic displacement charge is added to (6). This is only done in the case of the n-HACT structure at the lowest residual surface state concentration. At the other surface state concentrations in the n-HACT structure and at all surface state concentrations in the p-HACT structure the free surface boundary condition is replaced by a pinned fermi level boundary condition. The value of the fermi level is fixed to the value calculated from the one dimensional simulation with the fixed surface charge. The use of either the free surface with charge boundary condition or the pinned fermi level boundary condition in these cases results in the same potential profile. The similar potential profile occurs because the SAW displacement charge is a very minor perturbation compared to the fixed surface states, only changing the fermi level by $\sim 1.0 \times 10^{-5}$ volts. Even in the n-HACT structure at the lowest surface state concentration and a modified free surface condition of (6) with the acoustic displacement, the potential profile within the channel is unchanged from that obtained with a fixed fermi level at the surface. However, the pinned fermi level boundary condition tends to speed the convergence of the solution.

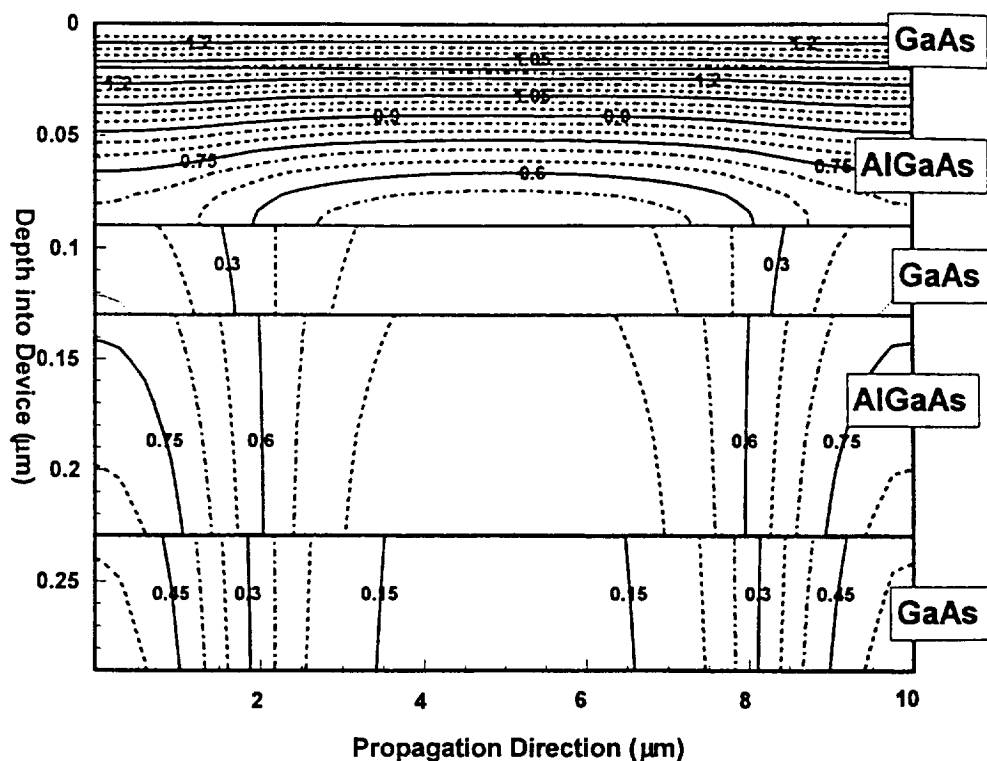
The simulation domain is nonuniformly discretized using rectangular control volumes. The node placement is performed adaptively during the equilibrium portion of the simulation. The adaptive gridding depends upon the change in potential, changes in material types or boundary conditions, and fluctuations in the doping concentrations. Generally the adaptive gridding results in a mesh consisting of approximately 107×41 points. The system of equations is linearized using Newton's method. The resulting system is solved using an iterative approach known as the conjugate gradient squared and stabilized method [9]. The code for this accelerator was added to the NSPCG library for the solution of nonsymmetric systems of equations [10].

B. Modeling the Surface Acoustic Wave

As previously stated the acoustic wave potential provides lateral confinement and transports the charge. To utilize the maximum coupling coefficient between the mechanical and electrical components of the SAW power a wave propagating in the $\langle 110 \rangle$ direction on a $[100]$ crystal structure is simulated. The SAW potential is calculated for the layer structures by the technique described in reference [11]. The resulting normalized potential as a function of distance from the surface for the p-HACT structure supporting a SAW with a power intensity of $1 \text{ mWatt}/\lambda$ is depicted in Fig. 6. The magnitude of the SAW potential for two different surface conditions, a free surface and a metallized surface, are displayed. Notice that the SAW potential is completely shorted out at the surface for the metallized device while the SAW potential maximum occurs at the surface in the free surface structure. Fig. 6 can be used to explain why a highly conductive surface can not be used in the HACT structure, as mentioned above. The location of both the n-HACT and p-HACT channels lie between the origin and the vertical line in Fig. 6. The channels are located very close to the surface to take advantage of the high magnitude of the



(a)



(b)

Fig. 9. Equi-energy contour plots of the n-HACT structure with an applied surface acoustic wave and three different residual surface state concentrations: (a) $-1.0 \times 10^{12} \text{ cm}^{-2}$, (b) $-1.5 \times 10^{12} \text{ cm}^{-2}$.

SAW potential present in the free surface case. As the front surface doping concentration increases, the surface becomes conductive and the maximum magnitude of the SAW potential

decreases in the immediate vicinity of the surface. Therefore, in the highly conductive surface device the SAW potential is insufficient near the surface to effectively transport charge.

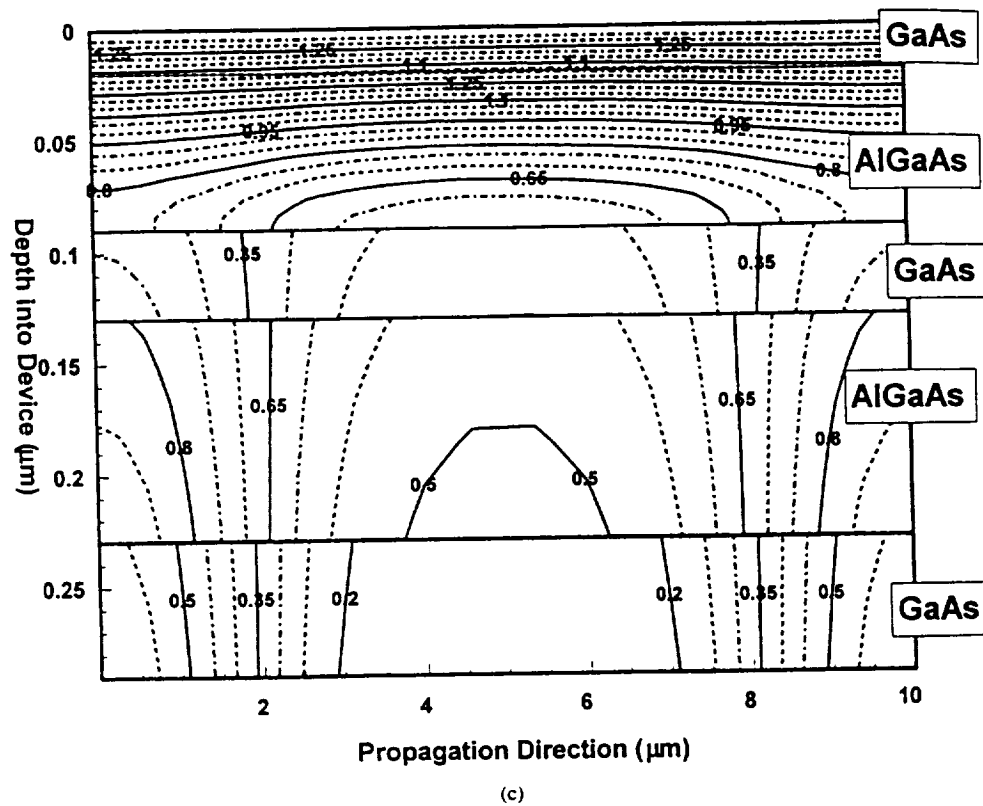


Fig. 9. (c) $-5.0 \times 10^{12} \text{ cm}^{-2}$.

As stated in Section III.B, the acoustic wave is modeled as a perturbation to the impurity doping density in Poisson's equation. The magnitude of the displacement charge resulting from the acoustic wave is obtained by the Laplacian of the wave potential and modulated by a sinusoidally varying function [12]

$$\Gamma_{\text{saaw}}(x, y, t) = P_s \frac{\epsilon}{q} \left[k^2 \phi_{\text{saaw}}(x) - \frac{\partial^2 \phi_{\text{saaw}}(x)}{\partial x^2} \right] \cos(ky - \omega t) \quad (8)$$

The factor P_s in (8) is a power scaling factor, indicating the SAW potential at the surface under free surface conditions, while k is the wave number of the acoustic wave. Fig. 7 depicts the maximum displacement charge, $\Gamma_{\text{SAW}}(x, y, t)$, in the p-HACT structure for an acoustic wavelength of $10 \mu\text{m}$, a P_s of 4 [12], and a free surface boundary condition. Contrary to the smooth behavior of the displacement charge in the case of bulk ACT devices the displacement charge in the HACT case shows discontinuities due to the abrupt changes in the permittivity of the constituent material.

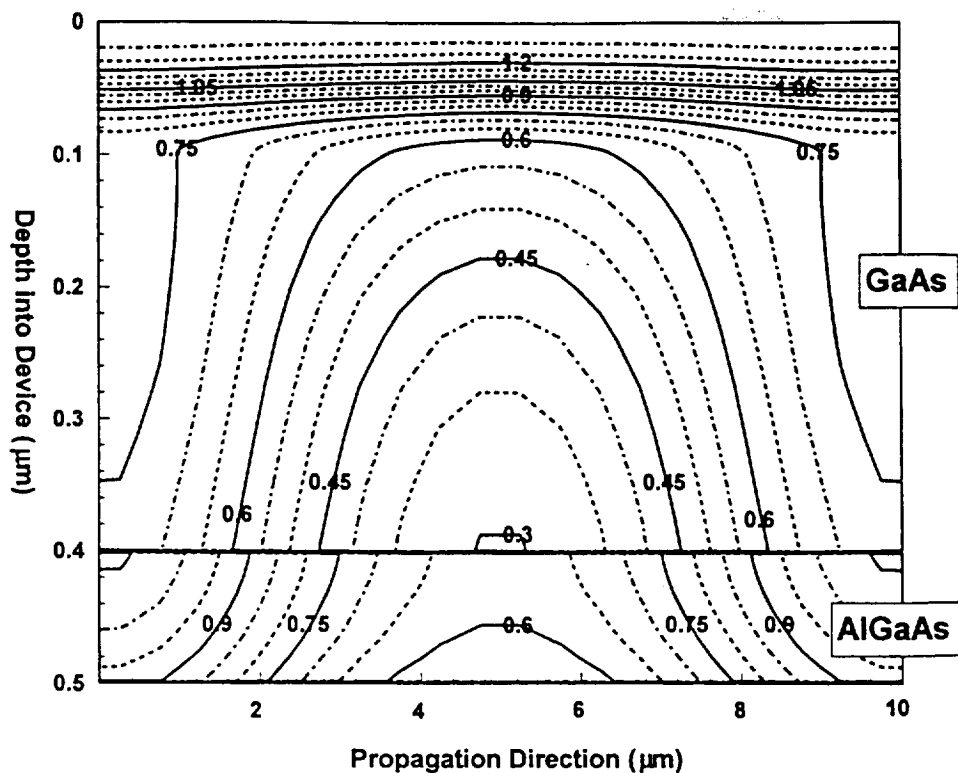
In the calculations presented below only one wavelength of the SAW will be simulated. The beginning and ending points are chosen at the potential minimums, thus creating a potential well to confine the electrons in the middle of the simulation domain. Provided the injected electrons do not greatly effect the shape of the SAW the choice of a symmetry boundary condition in the lateral directions can be justified.

C. Comparison with Previous Models

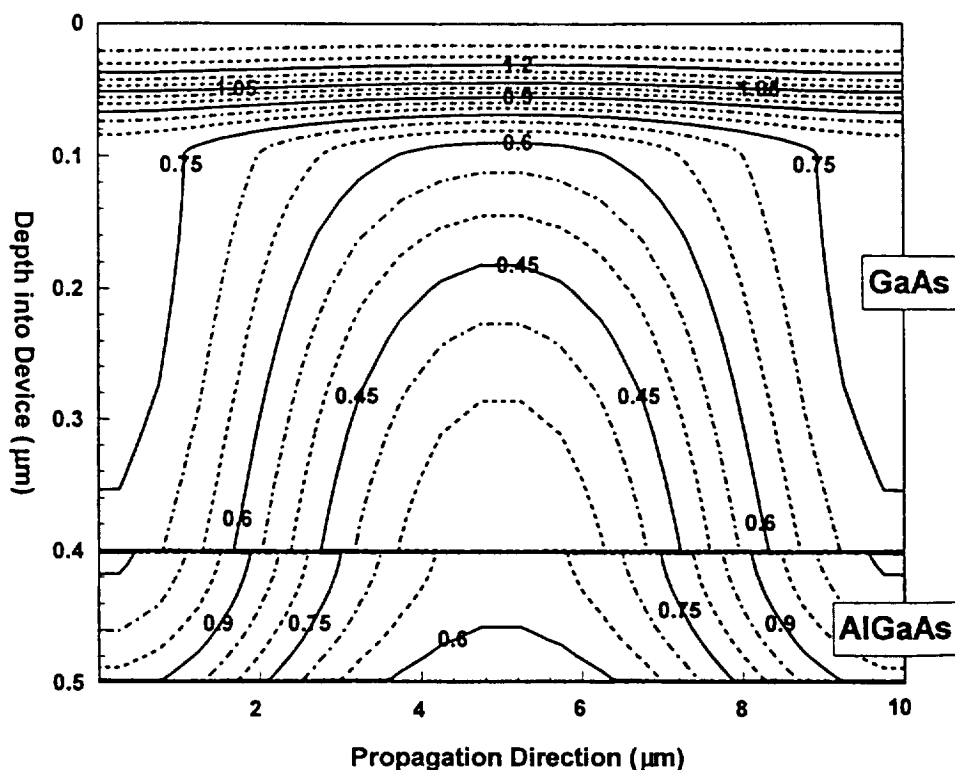
Several numerical models have been reported to simulate charge injection in bulk ACT devices [12]–[15]. One of the

major aims of this work is to analyze the HACT structures so only a limited comparison of our model to previous ACT models is appropriate. The earlier bulk ACT models neglect the hole contribution to the solution. This implies that no carrier recombination takes place, and therefore the charge transfer efficiency would always be calculated to be unity. Our model enables a more realistic calculation of the charge transfer efficiency by including the hole transport equation to simulate recombination. Previous ACT models were concerned with simulating the injection process between the input ohmic contact on the surface and the acoustic transport channel in the bulk of the device. In the model presented here the desired amount of charge is added at a specific location very close to the surface in a very small time step (1 psec). Then two small time steps are sequentially performed allowing the charge to distribute itself within the SAW's well. The time steps used in the simulation are 0.1 and 0.5 nsec. After 0.6 nsec elapses the injected charge is then in steady state with the acoustic potential. Next the time steps are increased to calculate the charge transfer efficiency.

Simulation of an n-HACT structure has been reported by researchers at United Technologies Research Center (UTRC) [2], [16]. As in the case of the bulk ACT models the hole equation was removed from the UTRC solution domain. The charge injection in the UTRC model is similar to the method of injection adopted here. A major difference between the two models is the UTRC model uses top and bottom ground planes as boundary conditions for the top and bottom of the device. It is in these planes that image charge created by the injected electrons will reside. Using the model presented here, which



(a)



(b)

Fig. 10. Equi-energy contour plots of the p-HACT structure with an applied surface acoustic wave and three different residual surface state concentrations: (a) $-5.0 \times 10^{11} \text{ cm}^{-2}$, (b) $-1.0 \times 10^{12} \text{ cm}^{-2}$.

does not use ground planes, it will be shown that the image charge actually resides in the substrate just below the bottom

heterojunction. Another major difference is in the calculation of the displacement charge. The UTRC model used a simple

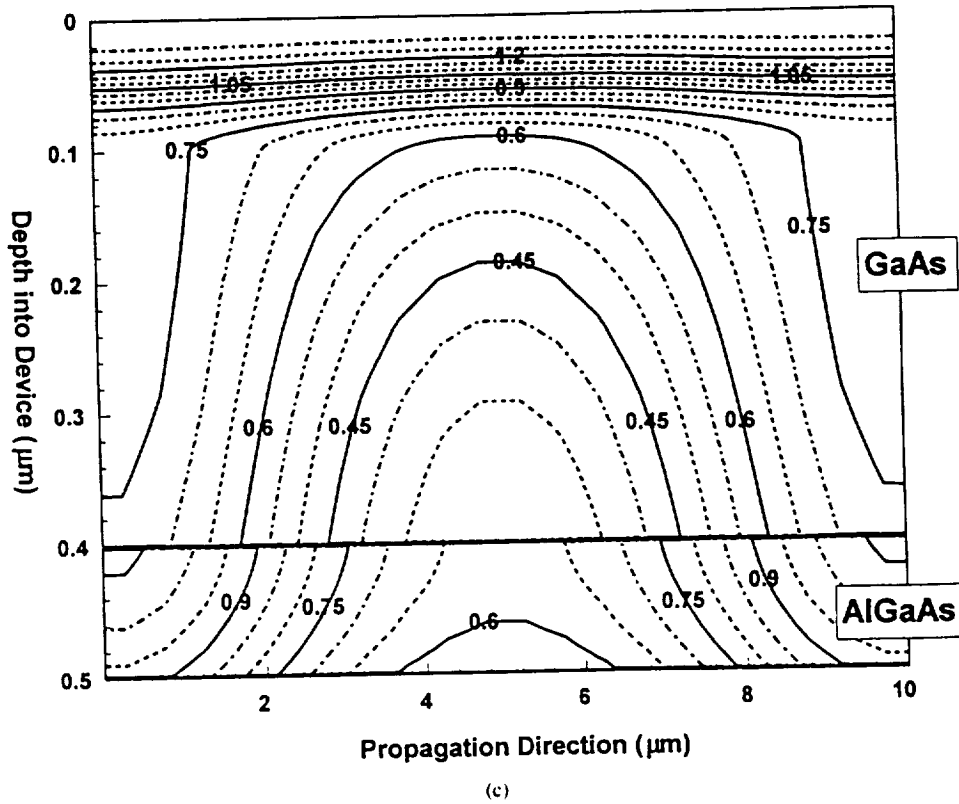


Fig. 10. (c) $-5.0 \times 10^{12} \text{ cm}^{-2}$.

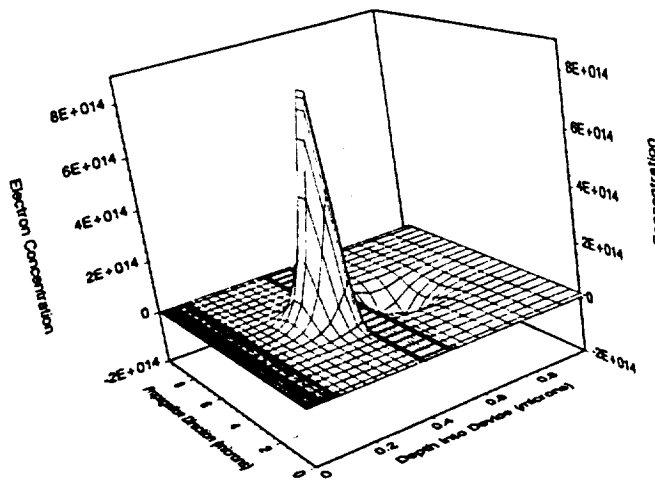


Fig. 11. Electron packet shape for the p-HACT structure with an applied surface acoustic wave and 10^6 electrons/cm injected into the transport channel. The loss of electrons from 0.6 to 0.8 μm deep is the image charge created by the injected electrons.

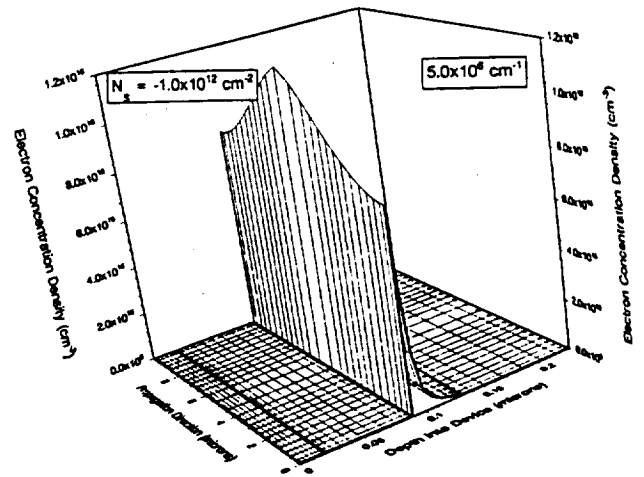


Fig. 12. Electron packet shape for the n-HACT structure with an applied surface acoustic wave, a residual surface state concentration of $-1.0 \times 10^{12} \text{ cm}^{-2}$, and an injected carrier concentration of 5×10^6 electrons/cm.

expression for the displacement charge in the HACT channel,

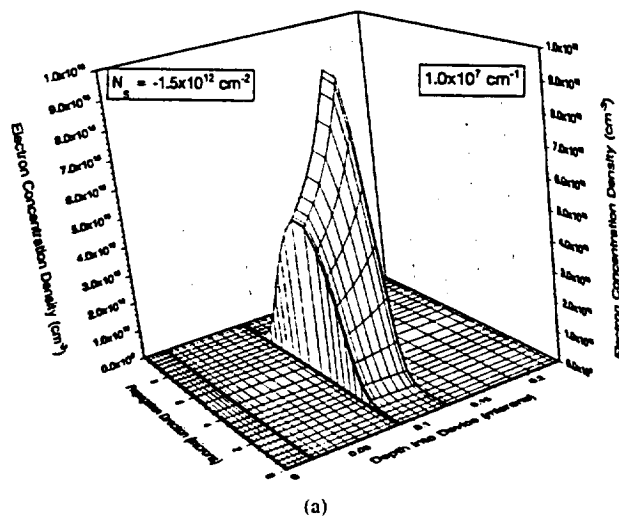
$$\Gamma_{\text{saaw}}(x, y, t) = -\left(\frac{\epsilon}{4\pi q}\right) \nabla^2(\phi_o \cos(ky - \omega t)) \quad (9)$$

where the constant ϕ_o was set to 1. In the model presented here, which uses (8), the value of the power parameter, P_s , would need to be set to an unusually high number as compared to that given in [12] to obtain the same potential profile as shown in [16]. The UTRC model and the model presented

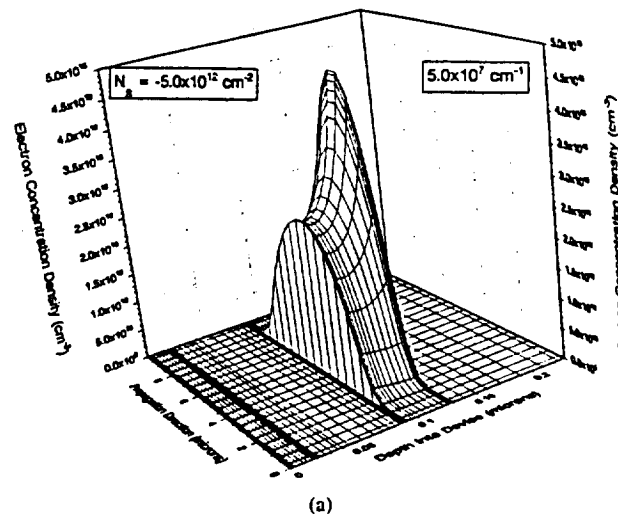
in this work both ignore the synchronous field created by the constant velocity of the SAW in the propagation direction [1].

IV. APPLICATION TO n-HACT AND p-HACT DEVICES

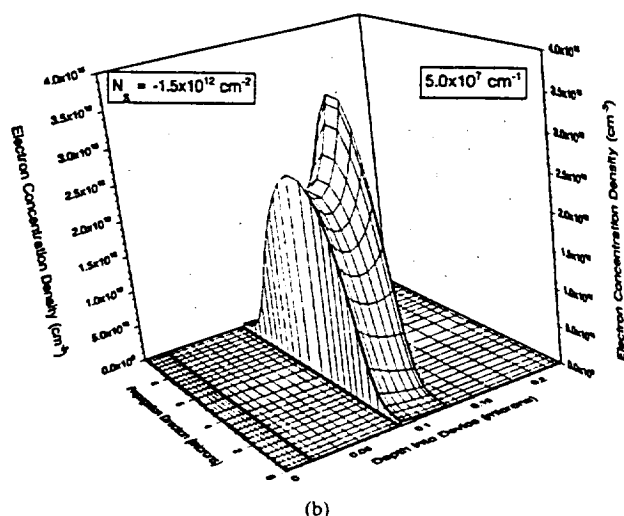
The one dimensional simulations, Figs. 3 and 5, depict the conduction band without any acoustic potential. These figures serve to highlight the effect of the surface state density for both the n-HACT and p-HACT structures. In this section these structures are simulated in two dimensions



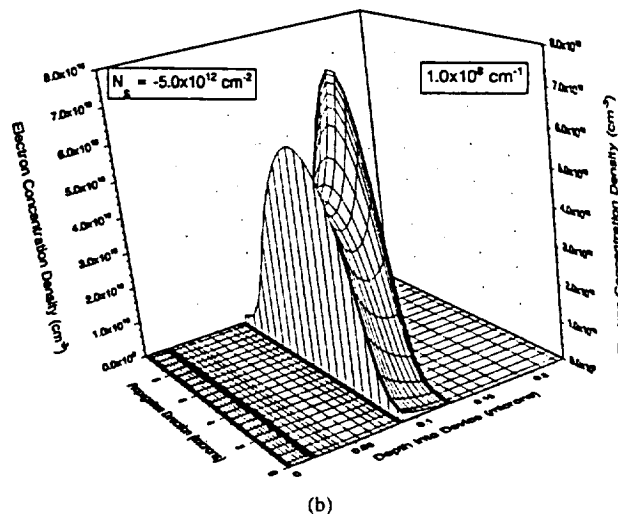
(a)



(a)



(b)



(b)

Fig. 13. Electron packet shape for the n-HACT structure with an applied surface acoustic wave, a residual surface state concentration of $-1.5 \times 10^{12} \text{ cm}^{-2}$, and two injected carrier concentrations: (a) 1.0×10^7 electrons/cm and (b) 5.0×10^7 electrons/cm. Notice the spreading in the case of 5.0×10^7 electrons/cm.

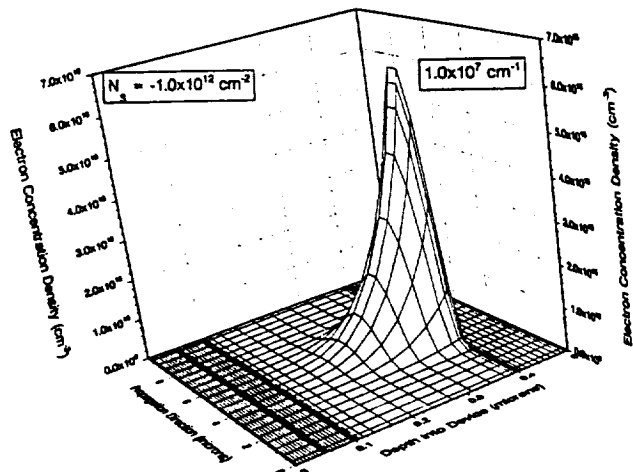
Fig. 14. Electron packet shape for the n-HACT structure with an applied surface acoustic wave, a residual surface state concentration of $-5.0 \times 10^{12} \text{ cm}^{-2}$, and two injected carrier concentrations: (a) 5.0×10^7 electrons/cm and (b) 1.0×10^8 electrons/cm.

with a superimposed SAW potential and injected charge to determine their charge capacity and charge transfer efficiency. The acoustic wave simulated has a wavelength of $10 \mu\text{m}$, a frequency of 287.75 MHz, which leads to an acoustic velocity of 2877.5 m/sec.

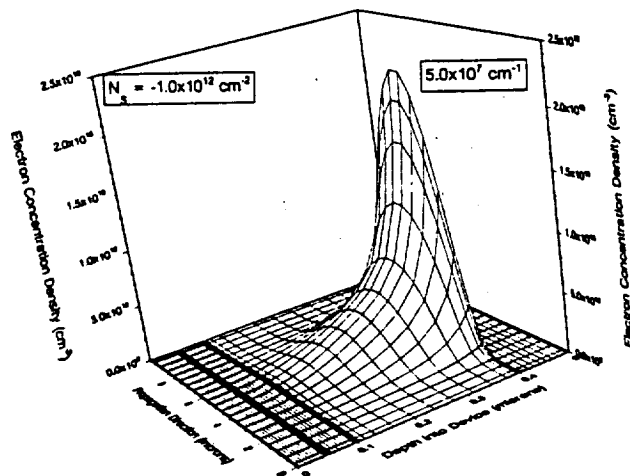
Fig. 8(a) and (b) show energy surface plots of the steady state conduction band for the n-HACT and p-HACT structures with a surface state density of $-1.5 \times 10^{12} \text{ cm}^{-2}$ and $-1.0 \times 10^{12} \text{ cm}^{-2}$ respectively. The power scaling factor, P_s , for both structures was 4. These figures are useful for illustrating the potential well created by the SAW potential. Hereafter, the conduction band will be illustrated by using contour plots of the energy surface. Comparison of Fig. 8(a) and (b) shows that the well created in the p-HACT structure (Fig. 8(b)) is approximately an order of magnitude larger than the n-HACT structure. In the n-HACT structure the top and bottom vertical confinement is provided by the two heterojunction barriers. As previously described, the top vertical confinement in the p-HACT structure is provided by the pn junction, with no

abrupt change in the conduction band, and the bottom vertical confinement is formed by the heterojunction discontinuity.

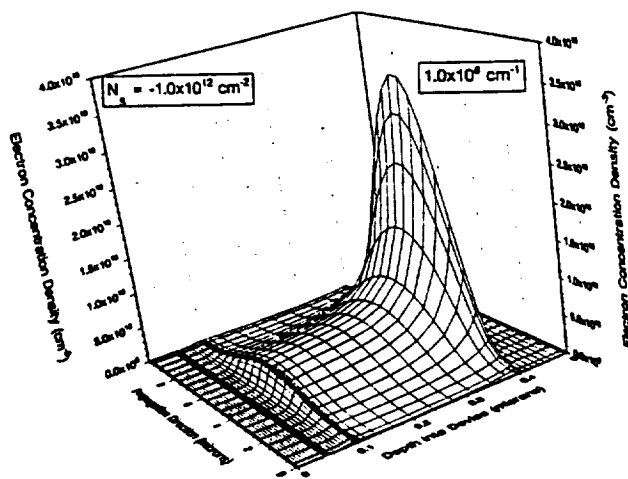
Fig. 9(a)–(c) show the equi-energy contour levels for the n-HACT structure with a P_s factor of 4 and surface state concentrations of -1×10^{12} , -1.5×10^{12} , and $-5.0 \times 10^{12} \text{ cm}^{-2}$. The values of the lateral barrier created by the SAW and the value of the pinned fermi level at the front surface are given in Table I. Note that there exists only a very little lateral potential barrier at a surface concentration of $-1.0 \times 10^{-12} \text{ cm}^{-2}$. The value of the pinned fermi level in Table I for the n-HACT structure demonstrates again this structure's sensitivity to the surface state density. At a residual concentration of $-1.0 \times 10^{-12} \text{ cm}^{-2}$ the level is 0.0895 eV, both the cap layer and channel layer are modulation doped by the AlGaAs. However, at residual concentration of $-1.5 \times 10^{-12} \text{ cm}^{-2}$ the level has dropped to -0.684 eV and the cap layer is now depleted. The steady state conduction band contours for the p-HACT structure with surface concentrations of -5×10^{11} , -1.0×10^{12} , and $-5.0 \times 10^{12} \text{ cm}^{-2}$ are displayed in Fig. 10(a)–(c). As in the case of the n-HACT structures the P_s value was



(a)



(b)



(c)

Fig. 15. Electron packet shape for the p-HACT structure with an applied surface acoustic wave, a residual surface state concentration of $-1.0 \times 10^{12} \text{ cm}^{-2}$, and three injected carrier concentrations: (a) 1.0×10^7 electrons/cm, (b) 5.0×10^7 electrons/cm, and (c) 1.0×10^8 electrons/cm. Notice the spreading in the case of 1.0×10^8 electrons/cm is much less than in the n-HACT structure Fig. 14.

TABLE I

n-HACT structure			
Surface States (cm^{-2})	-1.0×10^{12}	-1.5×10^{12}	-5.0×10^{12}
Pinned Fermi Level (eV)	0.0895	-0.6837	-0.8140
Lateral Barrier at Top of Channel (KT)	0.37	7.51	10.09
Lateral Barrier at Bottom of Channel (KT)	1.64	9.83	13.30
p-HACT structure			
Surface States (cm^{-2})	-5.0×10^{11}	-1.0×10^{12}	-5.0×10^{12}
Pinned Fermi Level (eV)	-0.6953	-0.7254	-0.8243
Lateral Barrier at Top of Channel (KT)	8.02	7.92	7.80
Lateral Barrier at Bottom of Channel (KT)	16.68	16.66	16.64

set to 4 and values of lateral barrier and pinned fermi level are listed in Table I. As seen from Fig. 10 and the values of the lateral barrier in Table I the surface concentration does not significantly change the shape of the potential well in the p-HACT structure.

To determine the maximum charge capacity of the potential well excess electrons are injected into the channel at the conduction band minimum. The structures are simulated until the electrons reach steady state with the acoustic potential. The number of injected electrons is increased until it appears that the electrons are not being confined within a single wavelength. Fig. 11 is an example of the electron charge packet shape when 10^6 electrons/cm are injected into the middle of the simulation domain for the p-HACT structure. To determine the electron packet shape the electron density simulated at steady state was subtracted from the electron density simulated 0.6 nanoseconds after the charge was injected. The 0.6 nanoseconds is required for the electrons which are only injected at the middle to attain a steady state condition with the acoustic potential. Notice that the electrons are confined to the center of the simulation domain, which is the lowest part of the conduction band in Fig. 10. The width of the packet is approximately $2 \mu\text{m}$ wide in the propagation direction. As will be seen in the figures to be presented the width of the packet increases as the amount of injected charge increases. Also note in Fig. 11 that an image charge, i.e. removal of electrons, has been created in the substrate just below the confining heterojunction. This image charge will not be plotted in later figures, but the reader should be aware that it exists.

Fig. 12 shows the electron packet shape for 5×10^6 electrons/cm injected into the n-HACT structure with the lowest residual surface state concentration. The electrons are not confined to the middle of the simulation domain, but spread over the channel along the propagation direction. This indicates that this structure has exceeded its maximum capacity for SAW transport; there is no confinement in the propagation direction. Fig. 13 depicts the electron packet shape for the n-HACT structure with a residual surface state concentration of $-1.5 \times 10^{12} \text{ cm}^{-2}$ for two injected electron concentrations. Fig. 13 a is for a packet of 10^7 electrons/cm. This is the maximum amount predicted

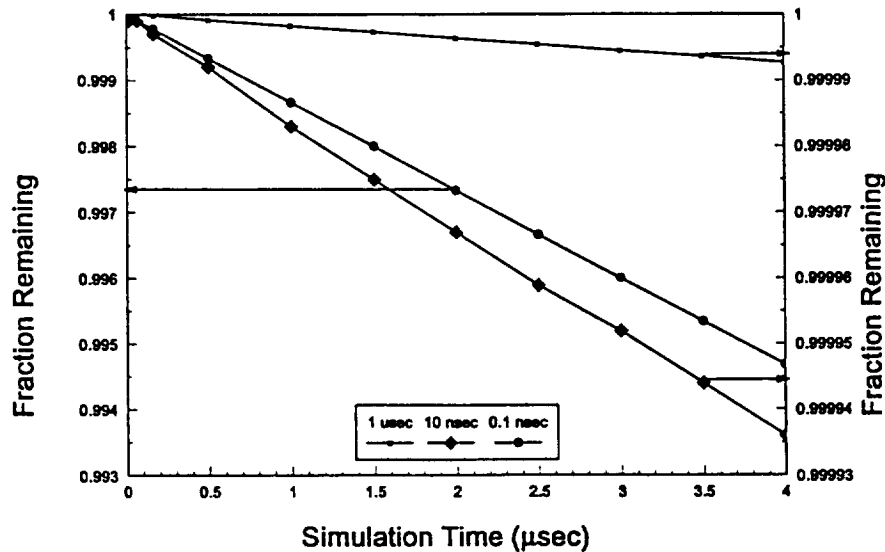


Fig. 16. Fraction of packet electrons remaining as a function of elapsed time in the p-HACT structure for three different values of SRH lifetime: 1 μsec, 10 nsec, and 0.1 nsec.

by the UTRC model for this structure. When the injected electron concentration is increased to 5×10^7 electrons/cm, Fig. 13(b), there is spreading of the packet over the entire simulation domain, indicating that the maximum capacity has been exceeded. Note that unlike the UTRC model the lack of confinement is due to the SAW potential not the heterojunction barrier. Fig. 14 displays the electron packet shape for the n-HACT structure with the highest residual surface state concentration. In this case the electron packet does not spread until the injected concentration reaches 10^8 electrons/cm. Comparison of Figs. 13 and 14 shows that excess surface charge can lead to a small increase in the maximum charge capacity of the wells.

The electron packet shape obtained using the p-HACT structure with a residual surface state concentration of $-1.0 \times 10^{12} \text{ cm}^{-2}$ and three different injected electron concentrations is shown in Fig. 15. For an injected electron concentration of 10^7 electrons/cm, Fig. 15(a), the packet is confined within $5 \mu\text{m}$ of the simulation domain. Fig. 15(b) shows the packet confinement when 5×10^7 electrons/cm are injected, even at 10^8 electrons/cm, Fig. 15(c), the packet is still fairly well confined to the simulation domain. Due to the similarities in the characteristics of the p-HACT structure at other residual surface state densities charge packet shapes at other surface state concentrations will not be presented. Simulations at surface state densities of 1.0×10^{11} and $1.0 \times 10^{13} \text{ cm}^{-2}$ also show similar insensitivity. Comparison of Figs. 13 and 15 shows that due to the larger channel thickness in the p-HACT structure the peak electron concentration density is approximately half of that obtained in the n-HACT structure for the same injected carrier concentration. These figures also show that for approximately the same surface concentration the p-HACT has a charge capacity which is nearly an order of magnitude higher than the n-HACT structure. The maximum charge capacity for the p-HACT structure is approximately the same as the n-HACT structure with a high residual surface

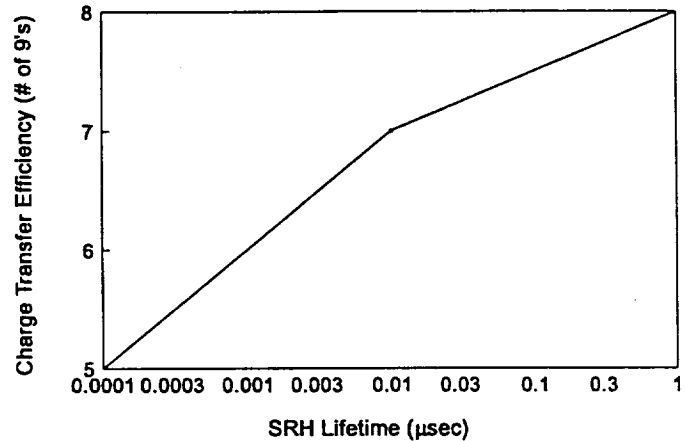


Fig. 17. Number of 9's in the charge transfer efficiency as a function of the value of the SRH lifetime in the channel region of the p-HACT device structure.

concentration, as can be seen from a comparison of Figs. 14 and 15.

To determine the maximum charge transfer efficiency of the p-HACT structure electrons are injected into the channel layer in the same manner as for the packet shape calculations. The structure is simulated for ~ 1000 wavelengths, $4 \mu\text{sec}$, at different time intervals. The amount of electrons remaining in the packet was recorded to determine the transfer efficiency. The fraction of electrons remaining in the packet as a function of time is shown in Fig. 16 for three different values of SRH lifetime in the channel layer. At high SRH lifetimes the fraction of remaining electrons is very close to unity, indicating excellent charge transfer efficiency. The calculated charge transfer efficiency, given by the number of 9's in efficiency, as a function of SRH lifetime is given in Fig. 17. For a SRH lifetime of 0.1 nsec the structure has 5 9's transfer efficiency and increases to 8 9's for 1 μsec lifetime. However, this analysis does not include any interface recombination at

the heterojunction nor does it account for slow traps in the channel layer. Both of these effects will act to lower the charge transfer efficiency.

V. CONCLUSIONS

An alternative HACT structure which uses a $pn^{-}np$ doping profile to deplete the transport layer and isolate the channel from the residual surface states has been analyzed. The $pn^{-}np$ doping also leads to an increase in the channel thickness. The p-HACT structure is insensitive to the residual surface state density over two orders of magnitude. Charge capacity simulations show that the p-HACT structure is capable of nearly an order of magnitude improvement in electron capacity over the n-HACT structure at nearly the same residual surface concentration. Simulations also show that the charge transfer is 5 to 8 9's efficient if the recombination is limited by SRH lifetimes in the range of 0.1 nsec to 1 μ sec. The p-HACT structure described in this paper has not been fully optimized, further calculations reveal that the doping in the first n-type layer may need to be reduced.

REFERENCES

- [1] R. L. Miller, C. E. Nothnick, and D. S. Bailey, "Acoustic Charge Transport: Device Technology and Applications," Artech House, Inc. Norwood, MA, 1992.
- [2] G. A. Peterson, B. J. McCartin, W. J. Tanski, and R. E. LaBarre, "Charge confinement in heterojunction acoustic charge transport devices," *Appl. Phys. Lett.*, vol. 55, pp. 1330-1332, 1989.
- [3] T. W. Grudkowski, R. N. Sacks, "HACT structure with reduced surface state effects," U.S. Patent #5264717, 1993.
- [4] A. W. Smith, "Light confinement and hydrodynamic modeling of semiconductor structures by volumetric techniques," Ph.D. Dissertation, Georgia Institute of Technology, 1992.
- [5] S. V. Pantankar, "Numerical Heat Transfer and Fluid Flow," Hemisphere Publishing Corp., New York, NY, 1980.
- [6] P. A. Basore, "PC-1D installation manual and user's guide version 3.0," Sandia National Laboratories, Albuquerque, NM, 1991.
- [7] K. Horio and H. Yanai, "Numerical modeling of heterojunctions using thermionic emission current at the heterojunction interface," in *Proc. of the 6th Inter. NASECODE Conf.*, pp. 390-395, Dublin, Ireland, July 11-14, 1989.
- [8] S. Selberherr, "Analysis and Simulation of Semiconductor Devices," Springer-Verlag, Wien, New York, 1984.
- [9] P. Chin and P. A. Forsyth, "A comparison of GMRES and CGSTAB accelerations for incompressible Navier-Stokes problems," *J. Comp. and Appl. Math.*, vol. 46, pp. 415-426, 1993.
- [10] T. C. Oppé, W. D. Joubert, and D. R. Kincaid, "An overview of NSPCG: a nonsymmetric preconditioned conjugate gradient package," *Computer Physics Communications*, vol. 53, pp. 283-293, 1989.
- [11] Y. Kim and W. D. Hunt, "Acoustic fields and velocities for surface-acoustic-wave propagation in multilayered structures: An extension of the Laguerre polynomial approach," *J. Appl. Phys.*, vol. 68, pp. 4993-4997, 1990.
- [12] E. G. Bogus, "Electrical charge injection in an acoustic charge transport device," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 1987.
- [13] E. G. Bogus, M. J. Hoskins, and B. J. Hunsinger, "Numerical model for electrical charge injection in the acoustic charge-transport device," *IEEE Trans. Electron Devices*, vol. 38, pp. 822-830, 1991.
- [14] S. M. Knapp, J. J. Liou, and D. C. Malocha, "Modeling the charge injection process in acoustic charge transport devices," *IEEE Ultrasonics Symp. Proc.*, pp. 223-227, 1989.
- [15] S. M. Knapp, D. C. Malocha and J. J. Liou, "A simplified and efficient numerical model for charge injection in acoustic charge transport devices," *IEEE Trans. Electron Devices*, vol. 39, pp. 1811-1820, 1992.
- [16] W. J. Tanski, D. E. Cullen, S. W. Merritt, R. N. Sacks, and G. A. Peterson, "Heterojunction acoustic charge transport devices for electronic warfare applications," Final Contractor Report to Air Force Aeronautical Laboratories #AFWAL-TR-88-1073, 1988.

Arlynn W. Smith received the B.S. degree in ceramic engineering from Alfred University, Hornell, NY, in 1984, the M.S. degree from the Georgia Institute of Technology, Atlanta, GA, in 1987, and the Ph.D. degree in electrical engineering from Georgia Tech in 1992.

From 1992 to 1993, he held a post-doctoral fellowship at the Georgia Tech Research Institute, where he developed a three-dimensional simulation code for modelling the chemical vapor infiltration of ceramic matrix composites. Currently, he is performing a post-doctoral research program with the microelectronics research center at Georgia Tech, developing a hydrodynamic simulation code of heterojunction acoustic charge transport devices, charge transport devices, and avalanche photodiodes.

J. Stevenson Kenney was born in St. Louis, MO, in 1962. He received the B.S. in electrical engineering in 1985, and the M.S. in electrical engineering in 1990, both from the Georgia Institute of Technology, Atlanta, GA. He is currently enrolled at Georgia Tech, pursuing the Ph.D. in electrical engineering.

He has over seven years of industry experience in microwave circuit and subsystem design, and has been employed by Electromagnetic Sciences, Inc., Scientific Atlanta and SPC Electronics America, Inc., all of Norcross, GA. He is currently employed by Pacific Monolithics, Inc., Sunnyvale, CA.

He has served on the steering for the 1993 International Microwave Symposium and received the Best Student Paper at the 1993 IEEE MTT-S International Microwave Symposium.



William D. Hunt was born in Jackson, MS, on December 21, 1954. He received the B.S. in 1976 from the University of Alabama, Tuscaloosa, AL, the M.S. from the Massachusetts Institute of Technology, Cambridge, MA, in 1980 and the Ph.D. from the University of Illinois at Champaign-Urbana in 1987, all in electrical engineering.

He worked for the Harris Corporation from 1976 to 1978, and for Bolt Beranek and Newman from 1980 to 1984. He joined the faculty of Georgia Tech in 1987, where he is currently an Associate

Professor. He was a Rhodes Scholar Finalist in 1975, received a DuPont Young Faculty Award in 1988 and was named an NSF Presidential Young Investigator in 1989. In addition, he has been selected for Who's Who in the South and Southwest and Who's Who in American Education and was named as one of the 1994 Distinguished Engineering Fellows by the University of Alabama College of Engineering.

His research interests include SAW and ACT devices, as well as transducers for biomedical ultrasound.



Kevin F. Brennan (S'84-M'84-SM'90) received the B.S. degree in physics from the Massachusetts Institute of Technology, Cambridge, MA, in 1978, and the M.S. degrees in physics and the Ph.D. degree in electrical engineering in 1980 and 1984, respectively, from the University of Illinois at Urbana-Champaign.

He is currently a Professor of Electrical and Computer Engineering at the Georgia Institute of Technology, Atlanta, GA. His current research interests include the theory and modeling of avalanche devices and confined state photomultipliers, the physics of impact ionization, nonlinear transport effects in semiconductors and heterostructures, plasmons in semiconductors, and the theory and modeling of acoustic charge transport devices.

Dr. Brennan was awarded a Presidential Young Investigator Award through the National Science Foundation in 1988.

Rudy Benz was born in Detroit, MI, on February 22, 1962. He received the B.S. degree with honors in engineering science from the Pennsylvania State University, University Park, PA, in 1984, and the M.S. and Ph.D. degrees in physics from the Georgia Institute of Technology, Atlanta, GA, in 1988 and 1992, respectively.

He is currently a research scientist at the Georgia Tech Research Institute. His research interests include the growth and characterization of II-VI and III-V semiconductors by molecular beam epitaxy, the surface kinetics of epitaxial growth, and the application of GaAs charge transfer devices to solid state imager applications.

Christopher J. Summers (M'82) was born in Oxford, England, on August 3, 1940. He received the B.S. and Ph.D. degrees in physics from Reading University, Reading, England, in 1962 and 1966, respectively.

After holding postdoctoral fellowship positions at Reading University and Bell telephone Laboratories, he joined GTE Laboratories in 1970 as a member of the technical staff. In 1972, he moved to McDonnell Douglas Research Laboratories, where he became a Research Staff Scientist. He joined the Georgia Tech Research Institute, Atlanta, GA, in 1981 and is currently a GTRI Fellow and Chief of the Quantum Microstructures Branch. His current research interests include optoelectronic properties of heterostructures, molecular beam epitaxy of II-IV and III-V semiconductors, and the growth and characterization of phosphor materials.

5.2: ZnO Films for ACT Devices

Here we present the work that we performed to ultimately reduce the RF power consumption of ACT devices. We have attached two papers which describes this work:

Kim, Y. and Hunt, W. D., "An analysis of surface acoustic wave propagation in a piezoelectric film over a GaAs/AlGaAs heterostructure," *Journal of Applied Physics*, vol. 71, no. 5, pp. 2136-2142, March 1992.

Kim, Y., Hunt, W.D., Hickernell, F.S., and Jen, C.K., "ZnO films on {001}-cut <110>-propagating GaAs substrates for surface acoustic wave device applications," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 42, no. 3, pp. 351-361, May 1995.

These papers indicate that ZnO is a suitable film choice for GaAs ACT devices because it can be deposited easily by sputtering and the SAW velocity for ZnO is comparable to that for GaAs. We also found out that ZnO can be made compatible with GaAs. Many GaAs processing people do not even like to hear the mention of Zinc because it is a fast diffusing, p-type dopant for GaAs. There concern is that if Zn enters any processing system that the GaAs wafers are going through, it will continue to dope Zn forever. This is probably a reasonable concern and in order for ZnO to be used it would have to be a final process step, after all other fabrication processing has been done. The wafer would need to leave the main facility and then go to another, separate location for ZnO deposition. We found that if the GaAs wafer has an encapsulation layer such as silicon nitride on it, the ZnO will not dope that wafer during sputtering. Hence, there is a way to get the ZnO layer on the GaAs on without harming the wafer. Though this may seem costly, there is a precedent for this in the silicon industry. Copper is now being hailed as a new metallization material for silicon and it too is a dopant for silicon. The interconnect layer that copper is used for is applied at the very end of the process and in a separate facility so that there can be no back contamination.

These two attached papers describe in considerable detail the work that was conducted. Ultimately, it was found that the ZnO film should be between 0.3 and 0.5 acoustic wavelengths thick for HACT devices.

An analysis of surface acoustic wave propagation in a piezoelectric film over a GaAs/AlGaAs heterostructure

Yoonkee Kim and William D. Hunt

School of Electrical Engineering and The Microelectronics Research Center, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 7 October 1991; accepted for publication 25 November 1991)

This paper presents a theoretical analysis of the surface acoustic wave (SAW) propagation in a GaAs/AlGaAs heterostructure covered with a piezoelectric film such as ZnO, BaTiO₃, PbTiO₃, or PZT. We calculate the SAW potential profile, SAW velocity, and effective piezoelectric coupling constant as functions of the film thickness using a Laguerre polynomial technique. The effective material constants such as the characteristic admittance and equivalent dielectric constant are also obtained. We concentrate on the suitable choice of the film for a heterojunction acoustic charge transport device. A ZnO film of 0.3–0.5 acoustic wavelength thickness seems to be the most suitable choice.

I. INTRODUCTION

Acoustic charge transport (ACT) devices have been developed for a variety of high-speed analog signal processing applications since they were first reported in 1982.¹ In ACT devices the charge packets are transported by the electric potential associated with a propagating surface acoustic wave (SAW). One ACT architecture¹ uses a thick (approximately half of the SAW wavelength) epitaxial layer and the charge is confined vertically by a combination of the depletion potential and the SAW potential. This epilayer of lightly doped GaAs causes difficulty in the monolithic integration of ACT devices with other electronic components. This problem has been overcome in part in a heterojunction ACT (HACT) device² in which the charge is confined in a GaAs well sandwiched between AlGaAs layers. The potential profile of a conventional HACT device is shown in Fig. 1 for both open- and short-circuit surface cases. For the open-circuit surface, the channel potential is the strongest because the ACT channel is located just underneath the surface. The potential is very small at the surface for the short-circuit case due to the electrical boundary conditions associated with a metal film.

For a signal processing application, the charge packet can be detected by nondestructive sense (NDS) electrodes. A complication in ACT devices which does not face the designer of charge coupled devices (CCD) is that the NDS array must perform its electrical function of sensing charge but must also not present an appreciable perturbation to the SAW. It has been shown that an electrode grating which has one eighth of an acoustic wavelength wide metallization and spacing, respectively, could be used to minimize the perturbation of the SAW by the NDS grating.³ As seen in Fig. 1, the presence of the metallic NDS array will short out the SAW potential at the surface, and this has been shown to adversely affect charge transport in the HACT structure. To overcome this problem with the HACT device, one needs an insulating layer between the ACT channel and the NDS electrodes. Therefore, the insulating layer helps to place the transport depth such that the channel potential for both the open- and short-circuit

cases is roughly equal. Moreover, if this insulating layer is a piezoelectric film, such as ZnO, it is possible to obtain a strong SAW potential and thus reduce the acoustic power necessary for charge transport. Considerable work on various film structures has been reported^{4,5} with most of the investigations being focused on the piezoelectric coupling coefficient, K^2 . For ACT devices, however, one wants to know not just the effective values of K^2 but the potential profile extending into the substrate.

This paper presents a theoretical analysis of the SAW propagation in a GaAs/AlGaAs HACT structure covered with a piezoelectric film. The numerical results are shown for film materials such as ZnO, BaTiO₃, PbTiO₃, or PZT. We use the Laguerre polynomial (LP) technique⁶ to describe SAW propagation in multilayered structures. Our results include the SAW velocities for the open- and short-circuit surface, SAW potential, and acoustic wave number. The value for K^2 is calculated from the relative change in velocities between the open- and short-circuit surface ($K^2 = 2\Delta V/V$).

II. METHODOLOGY

The schematic diagram of the structure to be investigated is illustrated in Fig. 2. The HACT device being considered consists of 400-Å-thick GaAs quantum well for charge transport sandwiched between the 700-Å-thick and 1-μm-thick AlGaAs layers. A 200-Å-thick GaAs layer is used as a cap layer. The mole fraction of AlAs in AlGaAs is chosen to be 0.35.

In our analysis, we used the stiffness, piezoelectric, and dielectric tensors reported by Adachi,⁷ Carlotti *et al.*,⁸ and Mansingh.⁵ These material tensors are transformed by the Euler angle method⁹ to reflect the desired crystal cut and propagation direction. As in the majority of ACT devices developed to date, the surface normal direction of the HACT device is (001) and the wave propagation direction is (110). The crystal symmetry of the ZnO film is hexagonal and that of the BaTiO₃, PbTiO₃, and PZT film is a poled ceramic. The polycrystalline ZnO film can be grown by various sputtering methods¹⁰ and its orientation is ob-

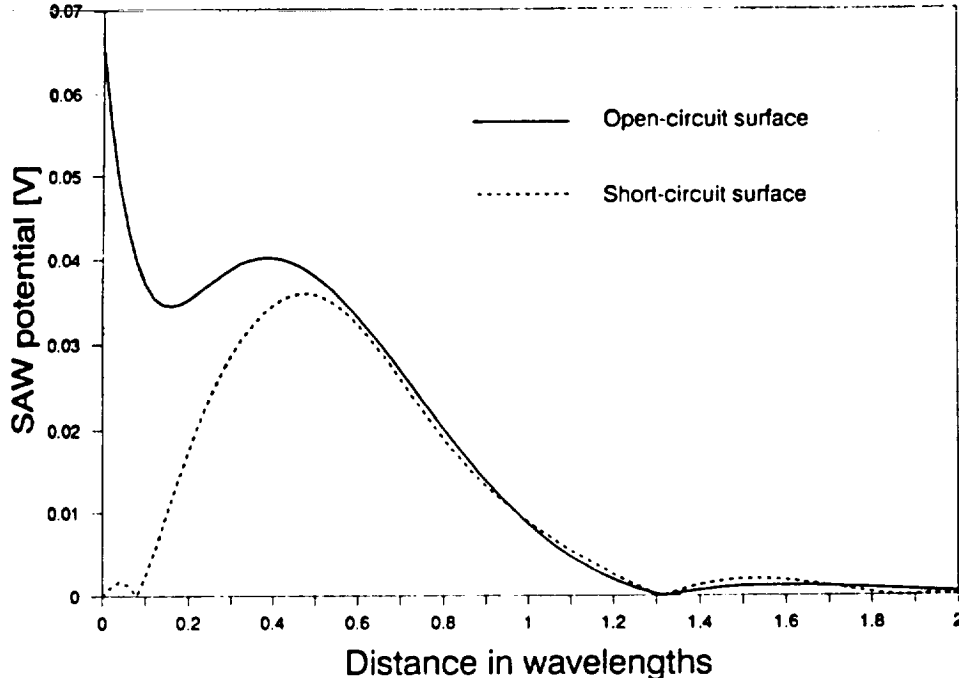


FIG. 1. SAW potential profile of a conventional HACT device for open- and short-circuit surface.

tained with the c axis normal to the surface by choosing proper sputtering parameters.¹¹ It should be noted that this property makes the poling process unnecessary for ZnO unlike for the piezoelectric ceramic films. It has also been reported that the c -oriented ZnO film can be deposited on top of a GaAs/AlGaAs/GaAs structure.¹² Since the basal plane of hexagonal crystal is isotropic, the propagation direction for calculation purposes is arbitrarily chosen to be (100). For the ceramic films, which are grown by various kinds of deposition methods,⁵ the material tensor has the same form as for the hexagonal crystal system, with Z axis along the poling axis. During the poling process, the poling axis is along the surface normal direction and it is parallel to the (001) direction.¹³ Therefore, the surface normal and the propagation directions are same as those of the sputtered ZnO film.

Using the LP technique, one can obtain the SAW potential profile in relation to the acoustic power. However, for ACT devices, it is preferable to have the SAW potential related to the voltage applied to the transducer. The equations presented in Datta¹⁴ are used in this paper. The sur-

face SAW potential, ϕ_0 , is given by $\phi_0 = \mu V_T$, where μ is the response function and V_T is the applied voltage to the interdigital transducer (IDT). The response function is given by

$$\mu = \mu_s(f_c)\eta N \frac{\sin[N\pi(f - f_c)/f_c]}{N\pi((f - f_c)/f_c)}, \quad (1)$$

where f_c is the center frequency of the IDT, N is the number of the IDT electrode pairs, and μ_s is the element factor which is equal to $0.8K^2$ for the metallization ratio $\eta = 0.5$. Since the SAW is a clock signal and is operated at the center frequency for ACT devices, μ becomes $0.8K^2$. Therefore, ϕ_0 can be simply written as

$$\phi_0 = 0.8K^2NV_T. \quad (2)$$

We make this value for ϕ_0 to be the surface SAW potential obtained by the LP technique.

The relation between the acoustic power, P_a , and ϕ_0 can be used for the calculation of the effective material constants. The characteristic admittance, y_0 , can be given by

$$P_a = |\phi_0|^2 \frac{y_0}{2} \frac{W}{\lambda}, \quad (3)$$

where W is the width of overlap of the IDT electrodes. The equivalent dielectric constant, c_s , is given by

$$c_s = K^2 \frac{y_0}{2\pi v_{\text{SAW}}} \quad (4)$$

where v_{SAW} is the SAW velocity.

III. RESULTS

The computed open-circuited SAW velocity versus film thickness is shown in Fig. 3. For our calculations, the acoustic wavelength, λ , of the filmed architecture is as-

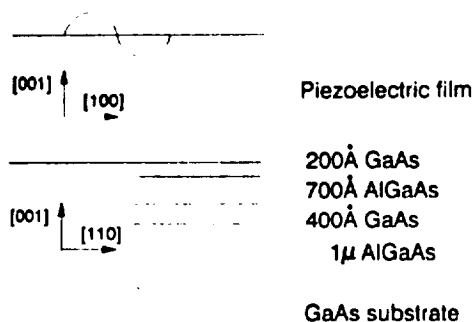


FIG. 2. Schematic diagram of a HACT device covered with a piezoelectric film.

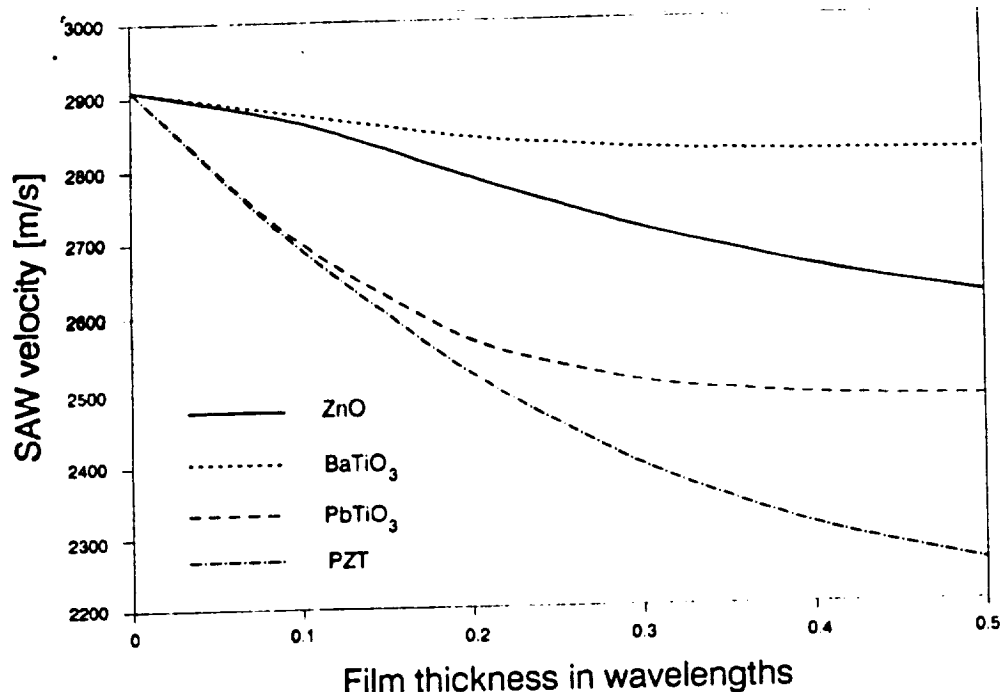


FIG. 3. Open-circuited SAW velocity vs film thickness.

sumed to be $10\text{ }\mu\text{m}$. We calculate over the range of $0\text{--}0.5\lambda$ thickness of the film. The Rayleigh modes are supported for all the filmed architectures being considered here. It is seen that the SAW velocity is close to that of the HACT substrate for the small thickness of the film and it is close to that of the corresponding single material as the thickness of the film is increased. Also note that the SAW velocity decreases monotonically. The SAW velocity of the HACT device which is drawn as the zero thickness of the film is calculated to be 2910 m/s (cf. $v_{\text{GaAs}} = 2864\text{ m/s}$). The variations of the SAW velocity by the 0.3λ thickness of the film are -6.6% , -2.9% , -13.9% , and -17.7% from that of the HACT device with no film overlay for the

ZnO, BaTiO₃, PbTiO₃, and PZT film, respectively. Note that the SAW velocity is changed only a little for the BaTiO₃ and ZnO film. Therefore it is expected that the bandwidth of the ACT device with these films is not decreased severely because the maximum bandwidth is half of the SAW clock frequency according to the Nyquist theorem ($f_c = v_{\text{SAW}}/\lambda$).

The value of K^2 versus film thickness is shown in Fig. 4. For the ZnO film, the value of K^2 increases rapidly up to the 0.1λ thick film, and reaches $0.8\%\text{--}0.9\%$ for the $0.1\text{--}0.5\lambda$ thick film. For the other films, the value of K^2 increases slightly for the $0\text{--}0.1\lambda$ thick film, and then increases rapidly up to the 0.3λ thick film, and reaches a peak at 2.41% , 1.50% , and 2.84% for the 0.5λ thick

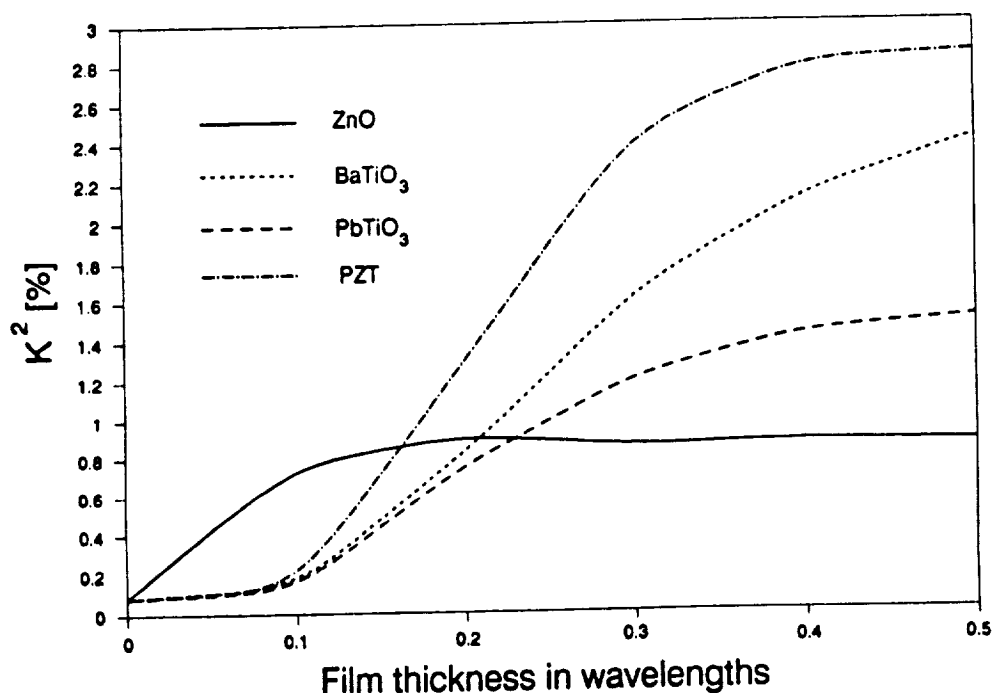


FIG. 4. Effective piezoelectric coupling constant vs film thickness.

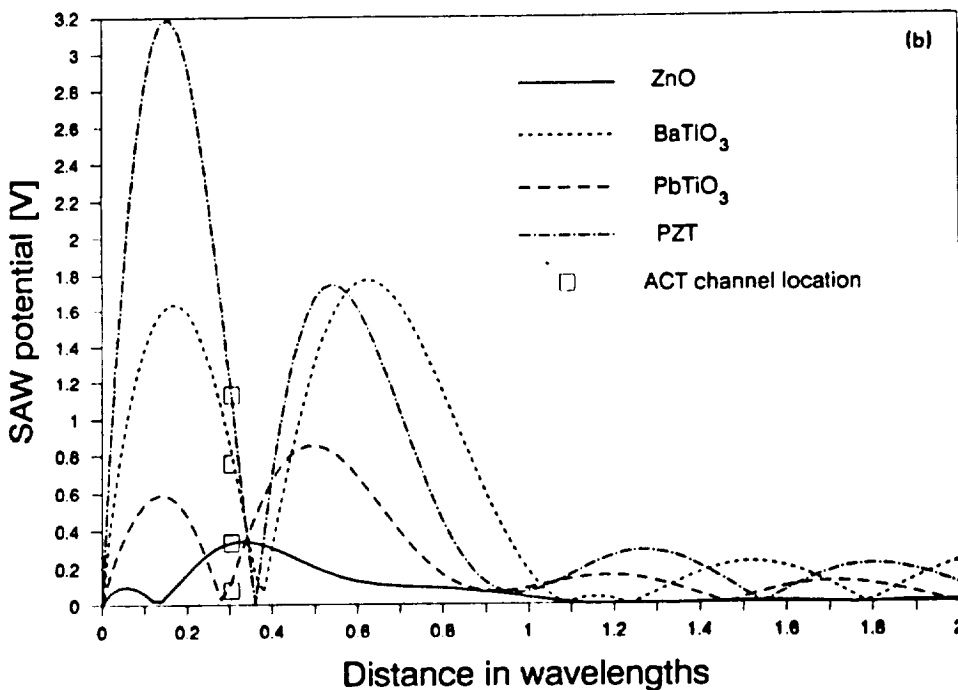
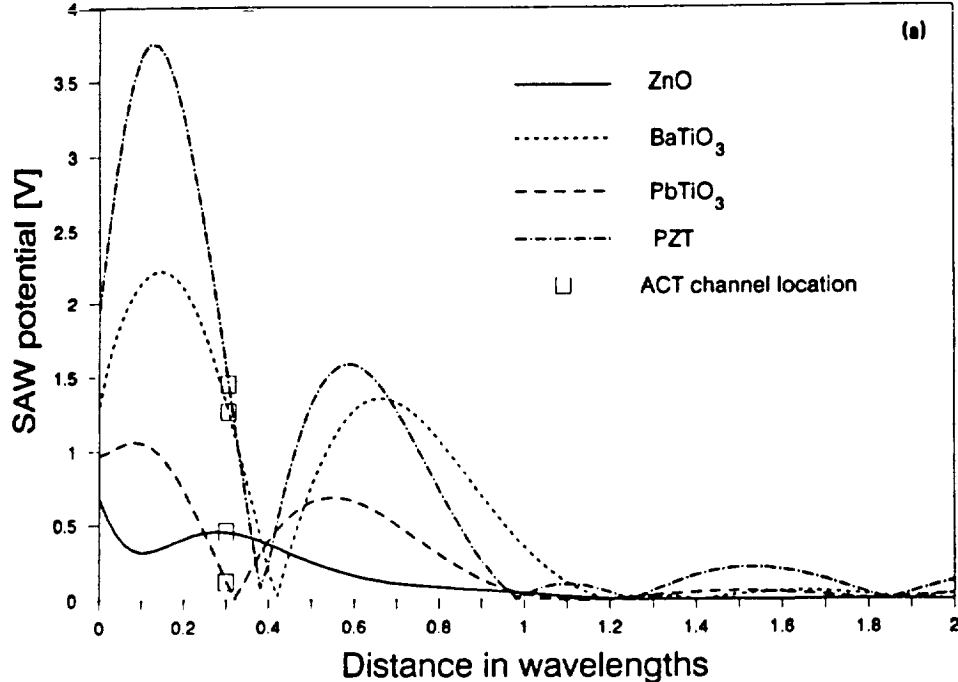


FIG. 5. SAW potential profile of 0.3λ thick film: (a) for open-circuit surface, (b) for short-circuit surface.

BaTiO₃, PbTiO₃, and PZT film, respectively. Note that the value of K^2 of the HACT device with no film overlay is calculated to be only 0.081%.

In order to compare the SAW potential, we assume an ideal power transfer condition in which the impedance of the IDT and that of the source are perfectly matched. We chose the values of N and V_T to be 100 finger pairs and 1V, respectively. Figures 5(a) and 5(b) show the calculated potential profiles of the films for the open- and short-circuit surface, respectively. The film thickness is chosen arbitrarily to be 0.3λ . Note that the phase shifting term is omitted to show the magnitude only. The location of the GaAs well for the charge transport channel is denoted by a square. It is seen that the channel potential is not pro-

portional to the value of K^2 especially for the PbTiO₃ film. Note that the local maximum point of the SAW potential is around the channel location for the ZnO film unlike for the other films. This characteristic is held for the range over the 0.3λ film thickness and it may help to confine the charge in the ACT channel not only horizontally but also vertically.

The calculated channel potentials versus the film thickness are shown in Figs. 6(a) and 6(b) for the open- and short-circuit surface, respectively. The channel potential of the HACT device with no film overlay is drawn as that of the zero film thickness. It is seen that the channel potentials are increased dramatically as the film thickness increases except for the PbTiO₃ film. From Fig. 6(b), it is

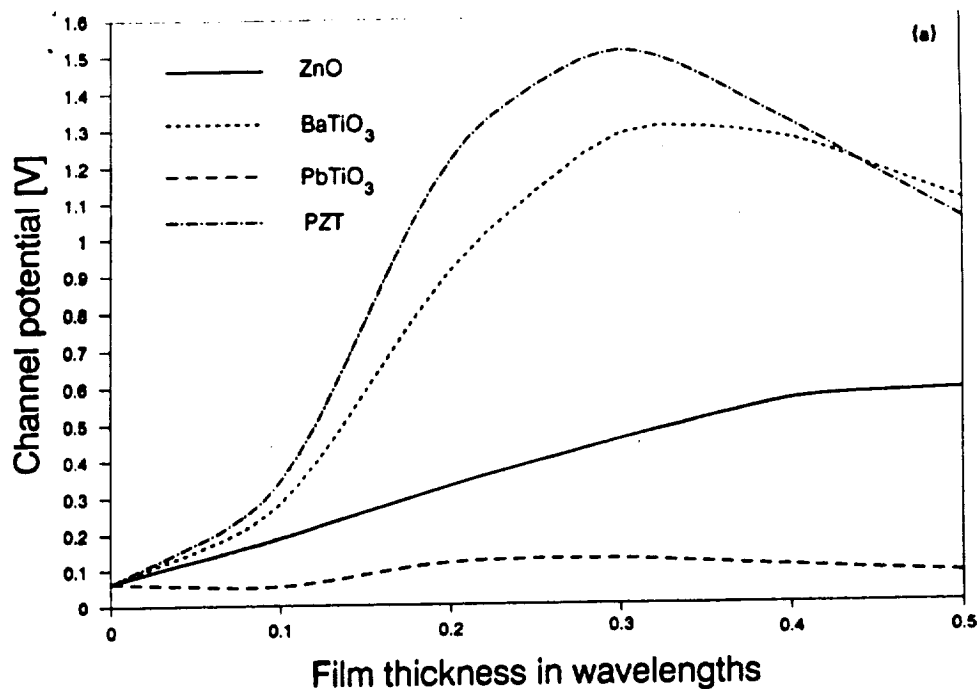
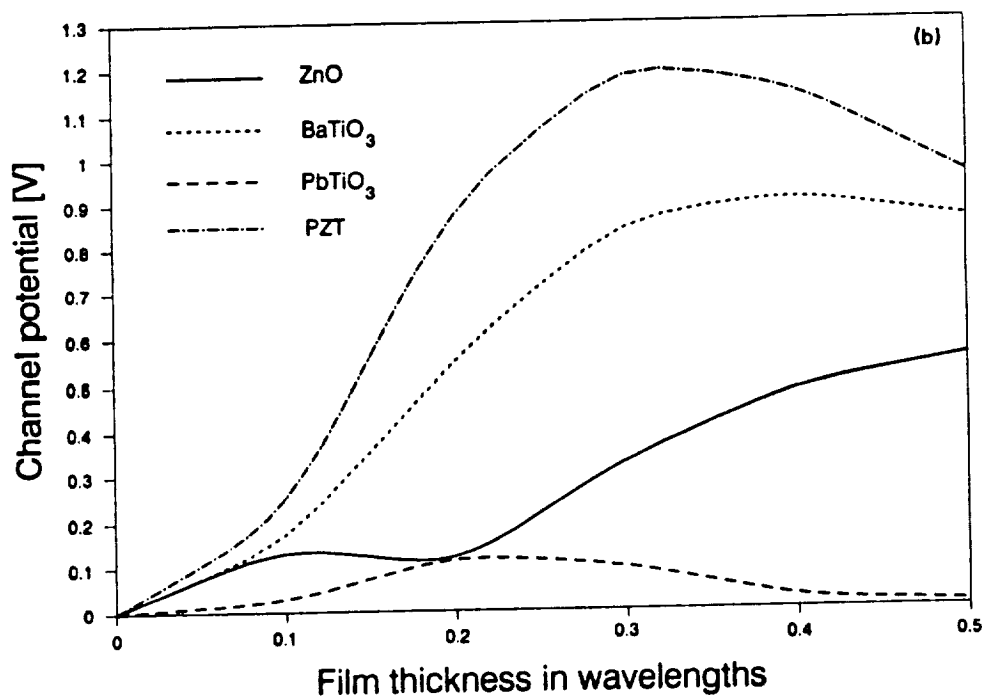


FIG. 6. Channel potential vs film thickness: (a) for open-circuit surface, (b) for short-circuit surface.



clear that the channel potentials for the short-circuit surface are not shorted out for the filmed structures while it is shorted out for the HACT architecture with no film overlay. This shows that the presence of the complicated NDS array does not cause the severe degradation of the charge transport properties. Even though the largest potential is expected in the PZT case, the ZnO film will be the most suitable choice because it does not require poling. Thus, we will present more about the ZnO film. The detailed SAW potential profiles for the ZnO film are shown in Figs. 7(a) and 7(b) for the open- and short-circuit surface, respectively. The decrease of the channel potential around the 0.2λ thickness shown in Fig. 6(b) is explained by the profiles shown in Fig. 7(b). Considering this for both the

open- and short-circuited potential, we can expect that the $0.3\text{--}0.5\lambda$ thickness is the best choice for the ZnO film. For this range, the open-circuited channel potential is shown to increase the channel potential by a factor of 7–9 over what it would be for the HACT architecture, and the short-circuited one maintains the value of 73%–94% of the open-circuited one.

Consider the ideal equivalent circuit of an IDT shown in Fig. 8. One can obtain the values for the circuit elements using the values of y_0 , and c , in the following equations presented in Datta.¹⁴ These values can be used to design a matching network of the IDT. The radiation conductance, G_r is given by

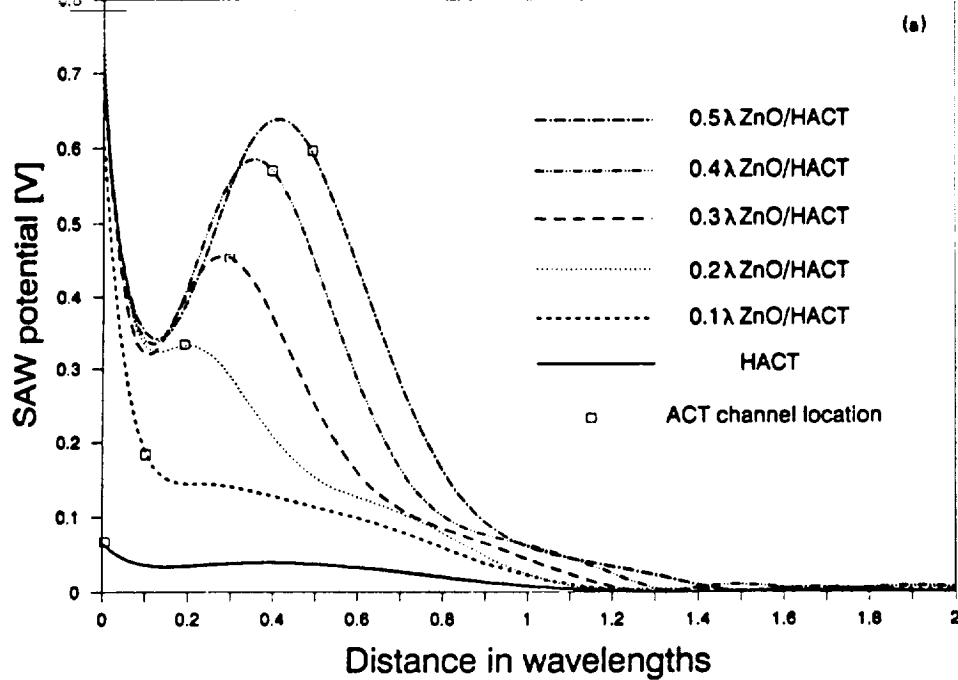
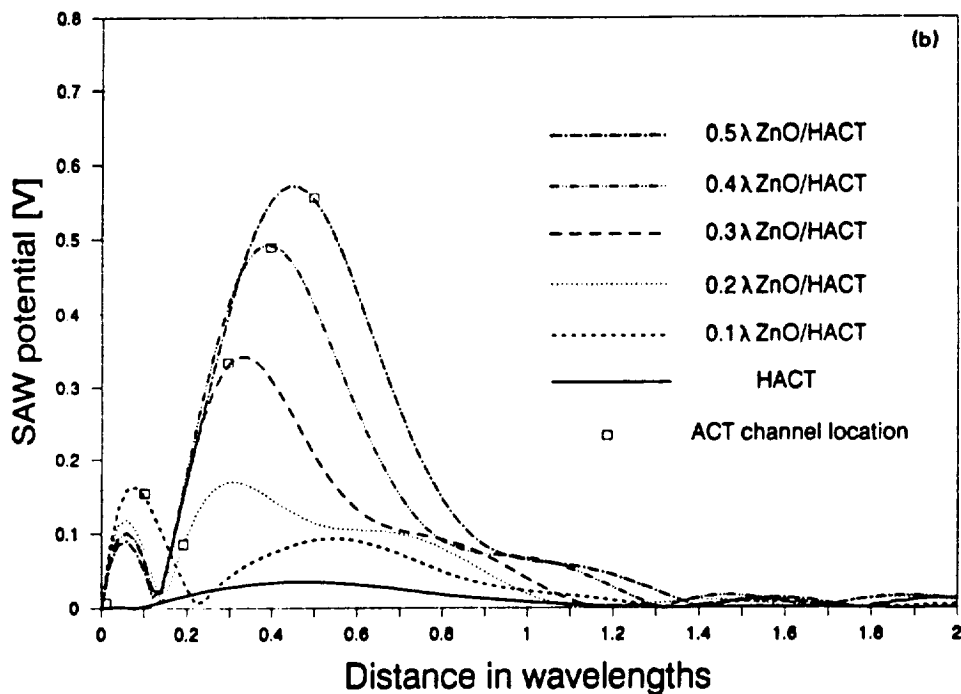


FIG. 7. SAW potential profile vs ZnO film thickness: (a) for open-circuit surface, (b) for short-circuit surface.



$$G_a = 2|\mu|^2 y_0 \frac{W}{\lambda}.$$

The radiation susceptance, B_a , is the Hilbert transform of G_a , and $B_a = 0$ at the center frequency. The equivalent capacitance, C_T , is given by

$$C_T = ac_s NW, \quad a = \begin{cases} 1 & \text{for solid-electrode IDT} \\ 1.4 & \text{split-electrode IDT} \end{cases}$$

The calculated y_0 and c_s of the HACT architecture covered with the ZnO film are shown in Fig. 9. The values for GaAs (Ref. 14) are given here for comparison,

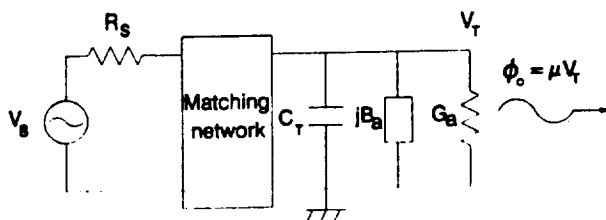


FIG. 8. Ideal equivalent circuit model of an IDT.

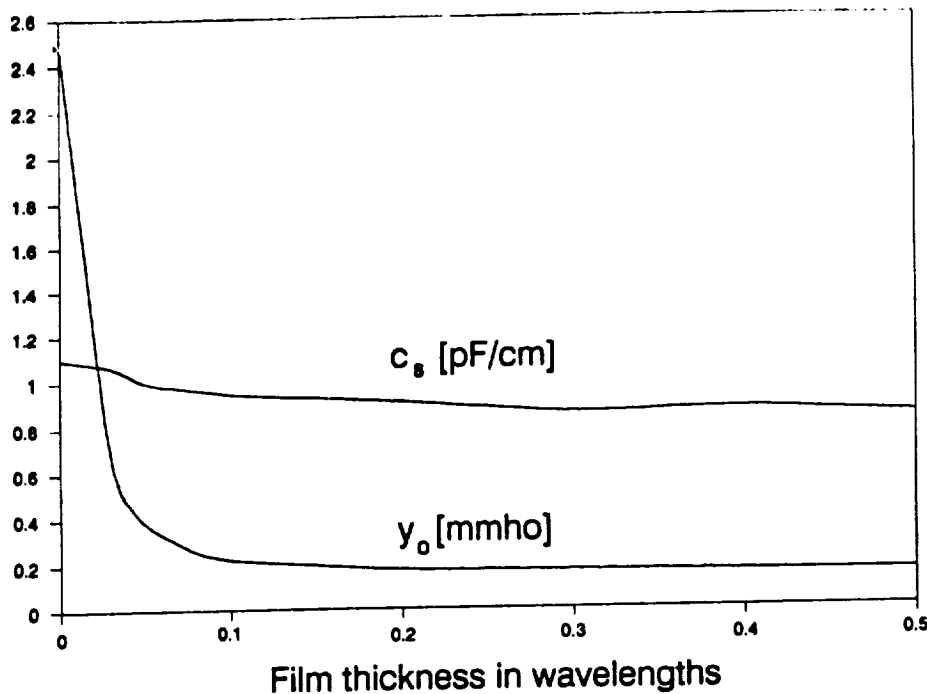


FIG. 9. Calculated characteristic admittance, y_0 , and equivalent dielectric constant, c_s , vs ZnO film thickness.

$$y_{0\text{GaAs}} = 3.1 \text{ mmho}$$

$$c_{s\text{GaAs}} = 1.2 \text{ pF/cm.}$$

IV. CONCLUSIONS

SAW velocity, K^2 , and potential profile of the piezoelectric films such as ZnO, BaTiO₃, PbTiO₃, or PZT on top of a GaAs/AlGaAs HACT device have been investigated using the Laguerre polynomial technique. Among these films, the ZnO film would be the most suitable choice. The effective characteristic admittance and dielectric constant for the HACT architecture covered with the ZnO film have also been calculated. It is found that the ZnO film of 0.3–0.5 λ thickness increases the open-circuited channel potential by a factor of 7–9 over the conventional HACT device and it makes the short-circuited one maintain the value of 73%–94% of the open-circuited one.

ACKNOWLEDGMENT

This research was supported by the National Science Foundation under a President Young Investigator Award for Professor Hunt.

- ¹M. J. Hoskins and B. J. Hunsinger, Proc. IEEE Ultrason. Symp. 456, (1982).
- ²W. J. Tanski *et al.*, Appl. Phys. Lett. 52, 18 (1988).
- ³W. D. Hunt and B. J. Hunsinger, J. Appl. Phys. 64, 1027 (1988).
- ⁴G. W. Farnell, Proc. IEEE Ultrason. Symp. 803 (1977).
- ⁵A. Mansingh, Ferroelectrics 102, 69 (1990).
- ⁶Y. Kim and W. D. Hunt, J. Appl. Phys. 68, 4993 (1990).
- ⁷S. Adachi, J. Appl. Phys. 58, R1 (1985).
- ⁸G. Carlotti *et al.*, Proc. IEEE Ultrason. Symp. 295 (1987).
- ⁹H. Goldstein, *Classical Mechanics* (Addison-Wesley, MA, 1980).
- ¹⁰F. S. Hickernell, Proc. IEEE Ultrason. Symp. 785 (1980).
- ¹¹O. Yamazaki, T. Mitsuyu, and K. Wasa, IEEE Tr. Sonics Ultrason. SU-27, 6 (1980).
- ¹²Y. Abdelrazek and C. S. Tsai, Proc. IEEE Ultrason. Symp. 495 (1989).
- ¹³H. Adachi *et al.*, Jpn. J. Appl. Phys. Suppl. 24-1, 121 (1985).
- ¹⁴S. Datta, *Surface Acoustic Wave Devices* (Prentice-Hall, NJ, 1986).

ZnO Films on {001}-Cut $\langle 110 \rangle$ -Propagating GaAs Substrates for Surface Acoustic Wave Device Applications

Yoonkee Kim, William D. Hunt, Frederick S. Hickernell, *Fellow, IEEE*, Robert J. Higgins, and Cheng-Kuei Jen

Abstract—A potential application for piezoelectric films on GaAs substrates is the monolithic integration of surface acoustic wave (SAW) devices with GaAs electronics. Knowledge of the SAW properties of the layered structure is critical for the optimum and accurate design of such devices. The acoustic properties of ZnO films sputtered on {001}-cut $\langle 110 \rangle$ -propagating GaAs substrates are investigated in this article, including SAW velocity, effective piezoelectric coupling constant, propagation loss, diffraction, velocity surface, and reflectivity of shorted and open metallic gratings. The measurements of these essential SAW properties for the frequency range between 180 and 360 MHz have been performed using a knife-edge laser probe for film thicknesses over the range of 1.6–4 μm and with films of different grain sizes. The high quality of dc triode sputtered films was observed as evidenced by high K^2 and low attenuation. The measurements of the velocity surface, which directly affects the SAW diffraction, on the bare and metalized ZnO on SiO_2 or Si_3N_4 on {001}-cut GaAs samples are reported using two different techniques: 1) knife-edge laser probe, 2) line-focus-beam scanning acoustic microscope. It was found that near the $\langle 110 \rangle$ propagation direction, the focusing SAW property of the bare GaAs changes into a nonfocusing one for the layered structure, but a reversed phenomenon exists near the $\langle 100 \rangle$ direction. Furthermore, to some extent the diffraction of the substrate can be controlled with the film thickness. The reflectivity of shorted and open gratings are also analyzed and measured. Zero reflectivity is observed for a shorted grating. There is good agreement between the measured data and theoretical values.

I. INTRODUCTION

OVER the past two decades, ZnO has been the most frequently considered piezoelectric film for use for increasing the piezoelectric coupling in cases of a weakly- (e.g.,

GaAs) or nonpiezoelectric substrate [1]. The use of a ZnO layer with a GaAs substrate will enable one to monolithically integrate surface acoustic wave (SAW) devices with GaAs electronics. Furthermore, such a structure will have low dispersion due to the similarity of the SAW velocities of both materials [2]. In this paper, the feasibility of ZnO films deposited on GaAs substrates for SAW device applications is investigated along with a detailed discussion about the problems which occurred during the characterization of the ZnO/GaAs structures.

Knowledge of SAW properties of the filmed structure is crucial for the accurate design of SAW devices. Herein, the experimentally determined fundamental SAW properties are reported including velocity, effective piezoelectric coupling constant K^2 , attenuation for the 1.6–4 μm thicknesses of c-oriented ZnO film over semi-insulating {001}-cut $\langle 110 \rangle$ -propagating GaAs substrates in the frequency range of 180–360 MHz.

Another critical acoustic property for the design of SAW devices is diffraction of the surface acoustic wave. Diffraction in anisotropic media can generally be described by the slowness surface or the velocity surface. The velocity surfaces on free and metalized surfaces were measured using two independent techniques: 1) knife-edge laser probe, 2) line-focus-beam scanning acoustic microscope (LFBSAM). The results from both methods are compared with the theoretical results.

Metallic gratings are basic elements required for the construction of SAW devices. Analyzing the reflectivity and the velocity change due to metallic gratings has been studied since the invention of SAW devices; however, no research on gratings on multilayered structures has been reported. Datta and Hunsinger have analyzed the grating properties on a single crystal by separating into the piezoelectric shorting and the mechanical scattering due to the mass loading and the stress [3]. In this paper, an extension of their technique to a multilayered structure is presented. Experimental results for a shorted and an open grating are compared with theoretical calculations.

A passivation layer such as SiO_2 [4], Si_3N_4 [5], or SiON [6] between ZnO and GaAs might be required in order to: 1) passivate the structure and enhance the yield, and 2) prevent unwanted doping of GaAs by Zn during the sputtering processes. The application of such a passivation

Manuscript received May 10, 1994; revised October 5, 1994; accepted October 7, 1994. This research was supported by Bell Northern Research Ltd. (BNR), the National Science Foundation under a Presidential Young Investigator Award for Professor Hunt, and by NASA under grant # NAGW-2753.

Y. Kim was with the Microelectronics Research Center, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA. He is now with the U.S. Army Research Laboratory, AMSRL-EP-ME, Fort Monmouth, NJ 07703-5601 USA.

W. D. Hunt is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

F. S. Hickernell is with Motorola Inc., Government and Systems Technology Group, Scottsdale, AZ 85252 USA.

R. J. Higgins is with Motorola Inc., Ft. Lauderdale, FL 33322 USA.

C. K. Jen is with Industrial Materials Institute, National Research Council, Quebec, J4B 6Y4 Canada.

IEEE Log Number 9409926.

TABLE I
ZnO FILM DEPOSITION PARAMETERS

Sputtering Method	RF Magnetron	DC Triode
Target	ceramic 6.5 inch dia.	ceramic 4 inch dia.
Target to Substrate	1.5 inches	2.5 inches
Background Pressure	8 mTorr	3 mTorr
Substrate Temperature	350°C	250°
Gas	82% Ar 18% O ₂	90% Ar 10% O ₂
Power	400W	150W
Rate	4.6 $\mu\text{m/hr}$	1.4 $\mu\text{m/hr}$

layer was obtained with a plasma-enhanced chemical vapor deposition (PECVD) for SiO₂ or Si₃N₄. The acoustic parameter differences with and without the passivation layer are discussed. Theoretical calculations of SAW velocity and K^2 as a function of film thickness were performed using the Laguerre polynomial technique [7].

II. ZnO FILM GROWTH

c-axis oriented polycrystalline ZnO films were grown at Motorola using both rf magnetron and dc triode sputtering method on {001}-cut GaAs wafers which were grown by the liquid-encapsulated Czochralski (LEC) technique. The GaAs wafers, which were chemically and mechanically polished on both sides, were semi-insulating with resistivity $\geq 10^7 \Omega\text{-cm}$, and the surface normal direction was $[100] \pm 0.1^\circ$. The quality of the film strongly depends on the fabrication conditions. The typical deposition parameters used for the ZnO film are listed in Table I. The film thicknesses of 1.6, 2.8, and 4.0 μm have been chosen to cover the range of 0.1–0.5 of the acoustic wavelength, λ . As a passivation layer, a 0.1 μm PECVD SiO₂ layer was deposited for the rf magnetron sputtered ZnO films and a 0.2 μm PECVD Si₃N₄ layer was deposited for the dc triode sputtered films. The choice of a passivation layer was made solely on the equipment availability.

The as-grown ZnO films were transparent with a very smooth surface finish. The grain size of the ZnO films was 0.2 to 0.5 μm for the rf magnetron sputtered films, and no grain boundaries are visible under scanning electron microscope (SEM) examination for the dc triode sputtered films indicating the grain size was beyond the resolving power of the SEM. From X-ray diffraction and atomic force microscope (AFM) measurements, the grain size of the dc triode sputtered films is 30 to 50 nm. The problem with these particular dc triode sputtered films, however, is that a compressive stress is induced in the substrate at the time the film is grown. The induced stress is greater with the thicker films such that 2" wafers are warped, making it difficult to process IDT's. Thus, only the ZnO film thickness of 1.6 μm was investigated for the dc triode sputtered films. It has been previously reported that a decrease in the internal compressive stress results in a decrease in K^2 [8].

The orientation of the ZnO film was characterized using the X-ray diffraction technique. The diffraction patterns showed that all samples had strong *c*-axis texture; i.e., there were no ZnO peaks visible except those from the basal planes.

TABLE II
CONFIGURATION OF IDT'S. NUMBER OF FINGER PAIRS N IS 50.5 FOR ALL IDTS

No.	Type	Nominal Wavelength (λ_n) [μm]	Measure Center Frequency (f_c) [MHz]	Aperture (W _g)	
				[μm]	(λ_n)
1	split	16	175.0-182.2	80	5
2	solid	12	230.2-240.6	100	8.3
3	solid	10	273.5-287.4	100	10
4	solid	8	339.0-356.8	80	10

III. CHARACTERIZATION OF ACOUSTIC PROPERTIES

A. Velocity, K^2 , and Attenuation

A number of devices were fabricated using a standard lift-off process on 1.6, 2.8, and 4 μm thicknesses of the ZnO films. Employing one set of four IDT's with different center frequencies enables the investigation of frequency characteristics over the range of 180–360 MHz. The configuration of the IDT's is listed in Table II. The values of the SAW wavelength, λ , in the table denote nominal values. Each IDT is unapodized with metallization ratio (finger width to grating period) of 0.6. The apertures of the IDT's were chosen to facilitate the measurement of the slowness surface on bare GaAs substrates [9]. Considering the nonacoustic resistivity of the IDT's [10], the metal thickness was chosen to be 0.1 μm of Al/4% Cu alloy. For ease of fabrication, the smallest dimension for the devices was no less than 2 μm ; therefore, IDT no. 1 utilized split finger electrodes while the others utilized the single finger configuration. Acoustic absorber was applied at both ends of the device in order to prevent SAW reflections from the edges.

The value of λ is inferred from the periodicity of the IDT electrodes, and the SAW velocity, v_{SAW} , in the IDT region may be obtained to first order by the relation, $v_{\text{SAW}} = f_c \lambda$, where f_c represents the center frequency. The value of the nominal λ within the free ZnO/GaAs surface region is different and is measured directly by using a technique such as the laser probe system. The measured values of f_c are shown in Table II, and the variation of f_c is due to the fact that v_{SAW} is dispersive depending upon the ZnO film thickness as well as the passivation layer.

A knife-edge laser probe system was used to measure SAW properties. The use of the laser probe system and the scheme for the longitudinal and transverse scans has been described in [11] and [12] and will not be discussed here. In this study, the signal to noise ratio of the laser probe was typically 40–50 dB, and a 3 μm diameter laser spot was used. From several devices on each substrate, the best device was chosen for the experiments. The SAW velocity measured with the knife-edge laser probe is shown in Fig. 1 along with the theoretical values. The velocity dispersion is plotted as a function of wavelength-normalized film thickness. The theoretical values were calculated with the value of λ set to 10 μm . The material constants in the calculation were those given by Slobodnik [13] and Hickernell [5].

The maximum difference between the SAW velocities of the free surface (measured with the laser probe) and in the IDT

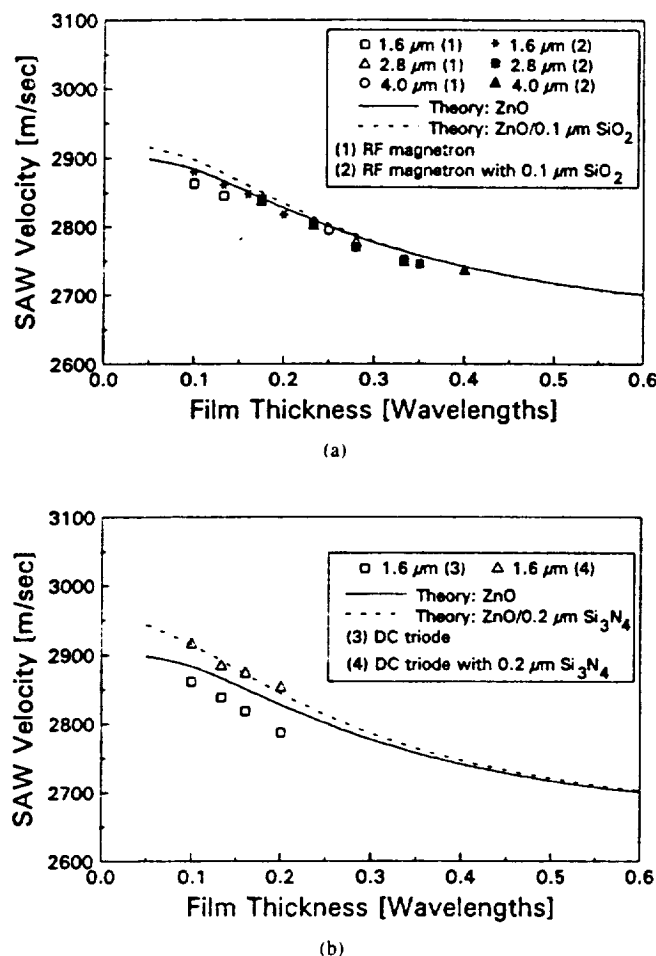


Fig. 1. SAW velocity measured by a knife-edge laser probe on the metalized surface: (a) RF magnetron sputtered film, (b) DC triode sputtered film.

area, from the center frequency measurements, was 0.3%. The maximum velocity deviations caused by the passivation layer were 0.6 and 1.6% for SiO₂ and Si₃N₄, respectively, and these figures agree well with the values from the center frequency measurements. The accuracy of the velocity measurement is approximately 0.3%. The detailed discussion about the accuracy can be found in Bright, Kim, and Hunt [14] and will not be reproduced here.

In using the laser probe, difficulty in making the free surface measurements was experienced due to the multiple reflections of the laser beam between the film-substrate interface and the air-film. For three film structures (2.8 and 4.0 μm rf magnetron sputtered ZnO and 1.6 μm dc triode sputtered ZnO/0.2 μm Si₃N₄), no appreciable energy from the laser beam was reflected from the free surface areas which is similar to what happens with antireflection coatings in optics. This made probing impossible. For this reason, the velocity shown in Fig. 1 is the measured data on the metalized surface as well as the theoretical value on short-circuit surface. More detailed analysis about this multiple reflection can be found in the previous article [15].

The effective piezoelectric coupling constant K^2 is generally obtained by a measurement of the relative shift in velocity

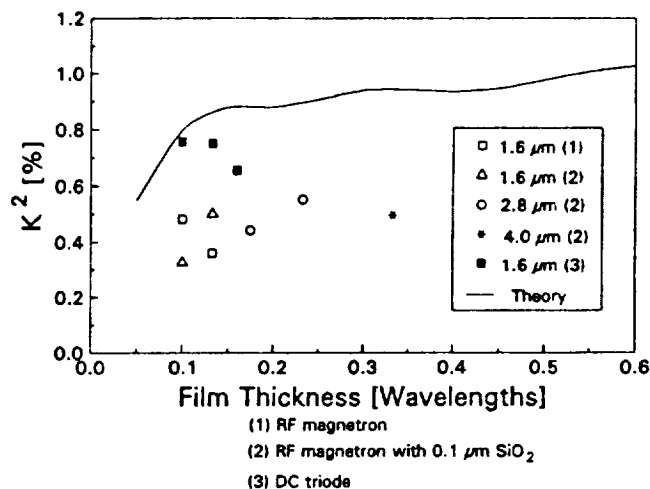


Fig. 2. Piezoelectric coupling constant K^2 .

between open- and massless short-circuit surface as follows,

$$K^2 = 2 \times \frac{v_o - v_s}{v_o} \quad (1)$$

where v_o and v_s are the open- and short-circuit velocity, respectively. For the measurement of v_s , a metal pad was fabricated on one side of the IDT's to short out the electric field on the surface while the other side of the IDT was left as a free surface. In order to reduce the mass loading of the metal pad, it was fabricated with 25 nm thickness of Al/4% Cu alloy. In spite of the finite thickness of the pad, its mass loading would be negligible because its thickness is only 0.0016–0.0031 of λ (16 to 8 μm) and it has a very slight effect compared to the piezoelectric effect [11]. This assumption was confirmed by a theoretical calculation using the Laguerre polynomial technique [7]. The velocity change due to the mass loading was calculated to be at most 0.015% which is much less than the expected value of K^2 for the ZnO/GaAs structure.

The value of K^2 calculated by using (1) is shown in Fig. 2 as a function of normalized film thickness. The data points represent average values obtained from several measurements. It is seen that the value of K^2 for the dc triode sputtered film is about 1.7 times larger than that of the rf magnetron sputtered films as expected from the grain size of the film. The measured value is 40–80% of the theoretical one, which is comparable to Hickernell's observations on the other substrates [1]. This value is still approximately 5.7–10.8 times larger than the typical value for bare GaAs (0.07%).

Acoustic propagation loss (attenuation) is also a crucial parameter in device applications of piezoelectric films. The attenuation of the SAW can be measured by comparing the energy contained in the wave of two different transverse scans, separated by some distance d . The energy is proportional to the value obtained by integrating the squared magnitude of the beam profile $|U|^2$ along the transverse direction. The scan length should be made long enough to include the entire energy since the beam spreads out due to diffraction. Thus,

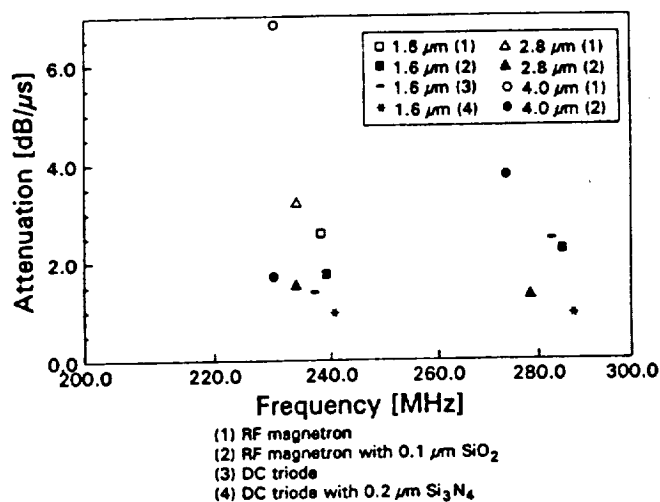


Fig. 3. Propagation loss in the frequency range of 230–290 MHz.

the attenuation α is given by

$$\alpha = 10 \log \left(\frac{\sum_1 |U|^2 \Delta x}{\sum_2 |U|^2 \Delta x} \right) \left(\frac{v_0}{d} \right) \text{ dB}/\mu\text{s} \quad (2)$$

where Δx is a step size for the scan. It should be noted that this attenuation includes not only 1) so-called viscous loss due to interaction with thermally excited elastic waves, 2) scattering by crystalline defects, impurities, and surface roughness, and 3) air loading but also the losses due to diffraction and the leaky nature of the wave on GaAs [16]. The effects of loss due to interaction with electrons should be negligible because of the high resistivity of the substrate.

The propagation loss shown in Fig. 3 was observed in the frequency range of 230–290 MHz. by comparing several transverse scans on the metalized surface with the distance of about 200λ and averaging the measured values. In fact, the diffraction was relatively large compared to that of bare GaAs due to the characteristic of the ZnO film to be discussed in the next section and the narrow aperture of the IDT ($W = 8.3$ or 10λ). With greater diffraction, more SAW energy can be lost to bulk shear since SAW propagation on this cut and propagation direction of GaAs is a leaky SAW mode. Employing an IDT with a wider aperture may reduce the value of the propagation loss in the measurement. Nevertheless, the narrow aperture was used in order to obtain enough diffraction for the velocity surface measurements to be discussed in the next section.

In the laser probe system, the output signal of the laser probe, which is detected with a photodiode, is filtered into dc and rf components. The dc component signifies the reflectivity of the sample under test, and the rf component is related to the normal component of the SAW particle displacement. While only the phase information of the rf component is required for the velocity measurement, its magnitude is used for the attenuation measurement. Since the magnitude of the rf component is proportional to the dc component, it is normalized by the dc value V_{dc} . The accuracy of the magnitude

measurement depends on, among other things, the flatness of the surface. However, the sample is not perfectly flat, resulting in a slight defocus of the laser probe spot and a variation in V_{dc} ; thus, precise measurement of attenuation is very difficult. Furthermore, due to the diffraction loss of the IDT, the maximum deviation of the attenuation measurement was up to a few tenths of a dB/ μ s depending on the scan distance.

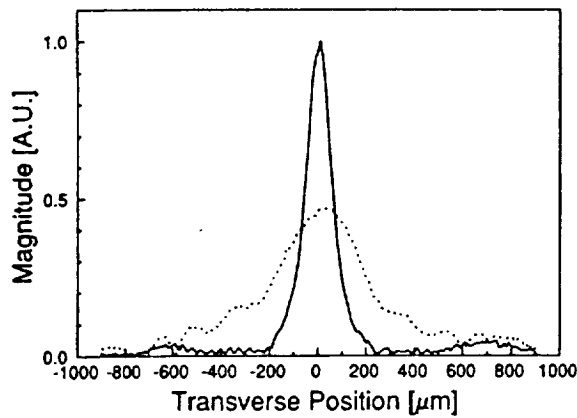
The measured data shows that the $1.6\mu\text{m}$ dc triode desputtered film with the passivation layer had the lowest loss and the $4\mu\text{m}$ rf magnetron sputtered film without the passivation layer the highest attenuation, which is comparable to what has previously been reported [1], [17]. The passivation layer had a noticeable effect on both sputtered films—especially for the rf magnetron sputtered ones. The reduction in the attenuation is perhaps due to the fact that the passivation layer provides a better platform for the film growth by compensating for stress differences between the film and the substrate [2]. In spite of the diffraction loss, it should be noted that the lowest value of attenuation for the films is not severely worse than that of bare GaAs substrate, which has the values of 0.323 and 0.567 dB/ μ s, for 200 and 300 MHz, respectively [18].

The ZnO/GaAs structures may not have suitable temperature stability required for some applications due to positive temperature coefficient delays (TCD's) of both bulk ZnO (35 ppm/ $^{\circ}\text{C}$) and GaAs (50 ppm/ $^{\circ}\text{C}$). However, since the passivation layer of SiO_2 (-70 ppm/ $^{\circ}\text{C}$) has a strong negative TCD, improvement of the temperature stability may be achieved by adjusting the passivation layer thickness [2]. An experimental compensation of TCD of GaAs was reported by using a Au/ SiO_2 film [19]. No measurements of the temperature characteristics have been attempted in this research.

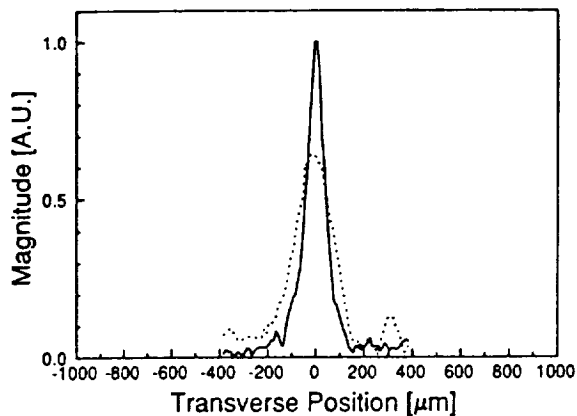
B. Velocity Surface

1) *Knife-Edge Laser Probe Measurements:* The substrates investigated for the velocity surface were 1.6 , 2.8 , and $4.0\mu\text{m}$ rf magnetron sputtered ZnO films with $0.1\mu\text{m}$ thick SiO_2 passivation layer and $1.6\mu\text{m}$ dc triode sputtered ones with and without $0.2\mu\text{m}$ thick Si_3N_4 passivation when λ was $12\mu\text{m}$. The typical transverse scan data measured on the free surface of $1.6\mu\text{m}$ dc triode sputtered ZnO/GaAs and bare GaAs are shown in Fig. 4(a) and (b), respectively, where the SAW propagation is along the $\langle 110 \rangle$ direction of GaAs. The two transverse scans were separated by a distance of 200λ . Note that the acoustic diffraction with the ZnO film shown in Fig. 4(a) is much larger than that of bare GaAs, which clearly shows a focusing beam profile in Fig. 4(b). Thus, the velocity surfaces of these structures are investigated theoretically and experimentally to explain the difference in the diffraction.

The theoretical velocity curves of the ZnO film with a $0.1\mu\text{m}$ SiO_2 passivation layer are shown in Fig. 5(a) as a function of the ZnO film thickness normalized to λ . The propagation directions are centered about the $\langle 110 \rangle$ axis. Note that the SAW velocity curves for the ZnO film thickness of 0.13 , 0.23 , and 0.33λ on GaAs are concave upward, and this is in contrast with that of the concave downward curvature for bare GaAs. The concave downward and upward curvature infer, respectively,



(a)



(b)

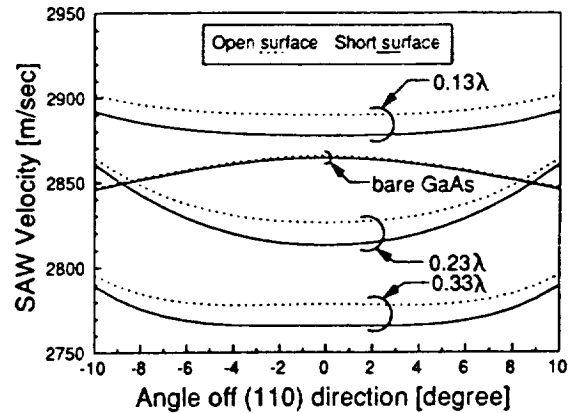
Fig. 4. Beam profiles are separated by a distance of 200λ when λ is $12\ \mu\text{m}$. The solid line represents the scan near the IDT and the dotted line represents that of 200λ apart. (a) Free surface of $1.6\ \mu\text{m}$ DC triode sputtered ZnO, (b) bare GaAs.

focusing and nonfocusing behavior, and support the diffraction shown in Fig. 4.

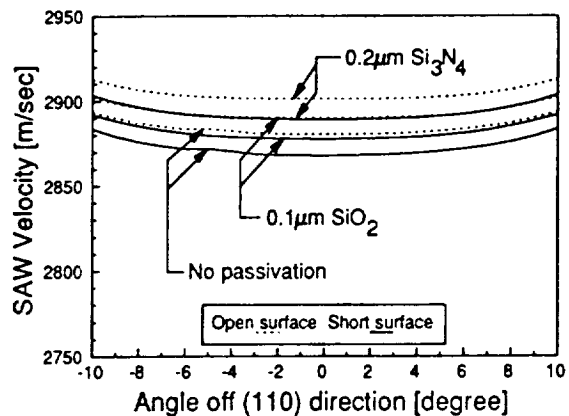
The effect of the passivation layer with $0.1\ \mu\text{m}$ thick SiO_2 or $0.2\ \mu\text{m}$ thick Si_3N_4 is shown along with nonpassivated structures in Fig. 5(b) for the ZnO thickness of $1.6\ \mu\text{m}$. Since the passivation layer is treated as an isotropic media like the c -oriented ZnO film, its presence is not expected to affect the shape of the curve, but shifts the entire curve slightly toward faster values. Considering the fast SAW velocities of single SiO_2 and Si_3N_4 , one can easily expect that the curve would shift upward as the thickness of the passivation layer increases as long as the energy is trapped at the surface.

Applying angular spectrum of plane waves (ASPW) theory to the transverse scans produced velocity surfaces shown in Fig. 6. The usage of the laser probe system for the ASPW theory has been described in detail in Hunt and Hunsinger [20]. Note that there is close agreement between Figs. 5 and 6.

The analysis of the velocity curve in terms of an anisotropy parameter gives an immediate insight into the diffraction of a propagating SAW. Using a parabolic approximation, the velocity of the pure mode direction ($\{110\}$ of GaAs) can be



(a)



(b)

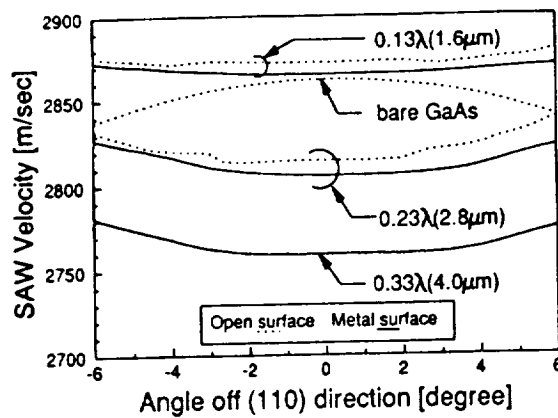
Fig. 5. Theoretical velocity surfaces obtained using Laguerre polynomial technique. Numbers indicate the thicknesses of the ZnO film normalized by $\lambda = 12\ \mu\text{m}$. (a) Effect of ZnO film thickness with a $0.1\ \mu\text{m}$ SiO_2 passivation layer, (b) Effect of passivation for 0.13λ thick ZnO film.

expressed as

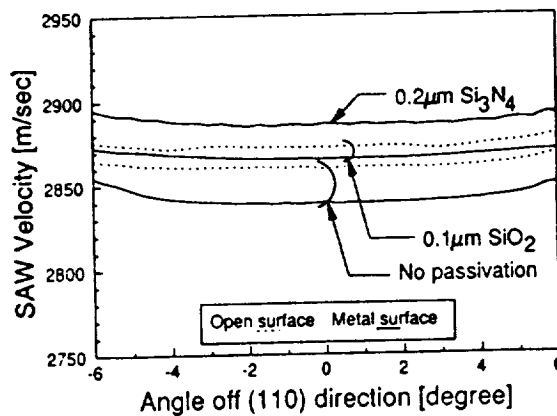
$$v(\theta) = v_0(1 - b\theta^2) \quad (3)$$

where v_0 is the velocity at $\theta = 0$, θ is an angle off the pure mode direction in radian, and b is the anisotropy parameter. The beam is spreading for $b < 0$ and focusing for $b > 0$ comparing to the isotropic case, $b = 0$ [21]. As shown in Fig. 5, the parabolic approximation for the 0.33λ thick ZnO film is not appropriate since the curvature is almost flat (i.e., isotropic) near the zero angle. Therefore, the values of b for the 0.13 and 0.23λ thickness were obtained from a least-square's curve fitting and are listed in Table III for the comparison between the theory and the experiment.

The inversion of the velocity curves was investigated numerically for the other materials. We arbitrarily chose GaAs, GaP, InAs, and InP since they are substrates with cubic symmetry, and AlN and ZnO as the piezoelectric film both of which have hexagonal symmetry. Material constants were available from Auld [22]. The inversion was observed for all of the combinations; e.g., it occurred at 0.05λ thick ZnO or 0.03λ thick AlN both on GaAs. The curves became flat as the film thickness was increased due to the isotropic characteristic of the films. In addition, SAW modes exist only for relatively



(a)



(b)

Fig. 6. Experimental velocity surfaces obtained applying ASPW theory to transverse scans. Numbers indicate the thicknesses of the ZnO film normalized by $\lambda = 12 \mu\text{m}$. (a) Effect of ZnO film thickness with a $0.1 \mu\text{m}$ SiO_2 passivation layer. (b) Effect of passivation for 0.13λ ($1.6 \mu\text{m}$) thick ZnO film.

thin films of AlN due to its much faster velocity (5608 m/s) than those of the substrates (1990–3540 m/s); e.g., for 0.11λ or less thickness of AlN on GaAs. For comparison, the velocity curve was also calculated for an isotropic film such as fused silica (3411 m/s) on GaAs, but in this case the inversion did not occur. The calculation results suggest that both the hexagonal films (AlN and ZnO) cause the inversion of the velocity surface in a cubic material for the cases mentioned above. This indicates that it is possible to control the diffraction of the substrate with proper selection of film thickness.

2) Line-Focus-Beam Scanning Acoustic Microscope Measurements: Again, Fig. 5(a) indicates that the velocity surface inversion occurs about the $\langle 110 \rangle$ direction when the c -oriented ZnO film is deposited on a $\{001\}$ -cut GaAs wafer. Our theoretical calculations also reveal that the reverse situation appears if the propagation is along another pure mode axis $\langle 100 \rangle$ direction. In order to evaluate this prediction and allow the measurement of acoustic properties along nonpiezoelectric directions, velocity measurements using a Line-Focus-Beam scanning acoustic microscope (LFBSAM) [23] operating at 225 MHz were performed. The technique to deduce the velocity is the so-called $V(z)$ analysis which has been reported

TABLE III
ANISOTROPIC FACTOR OF 0.13 AND 0.23 λ THICKNESS OF ZnO FILM. THE VALUES ARE OBTAINED BY CURVE FITTING WITH A LEAST-SQUARE METHOD. THE VALUES IN THE PARENTHESES ARE FOR THE SHORTED SURFACE

ZnO Thickness	Sputtering Method	Passivation	Theory	Experiment
0.0	—	0	0.22	0.45 ^a
0.13 λ	RF Magnetron	1	-0.13 (-0.16)	-0.18 (-0.17)
0.13 λ	DC Triode	2	-0.13 (-0.16)	N/A (-0.19)
0.13 λ	DC Triode	0	-0.14 (-0.17)	-0.22 (-0.4)
0.23 λ	RF Magnetron	1	-0.44 (-0.56)	-0.71 (-0.55)

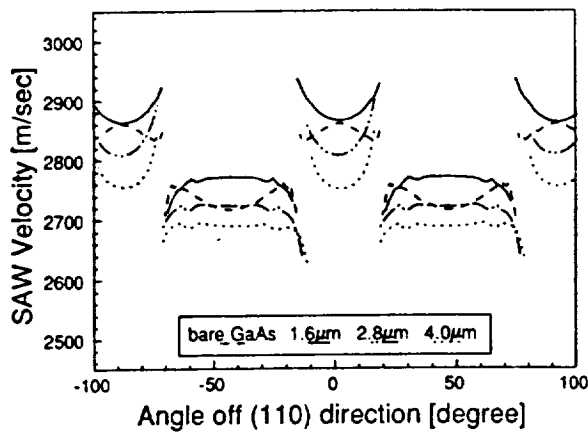
0: None, 1: $0.1 \mu\text{m}$ SiO_2 , 2: $0.2 \mu\text{m}$ Si_3N_4 .

^a Reference [11].

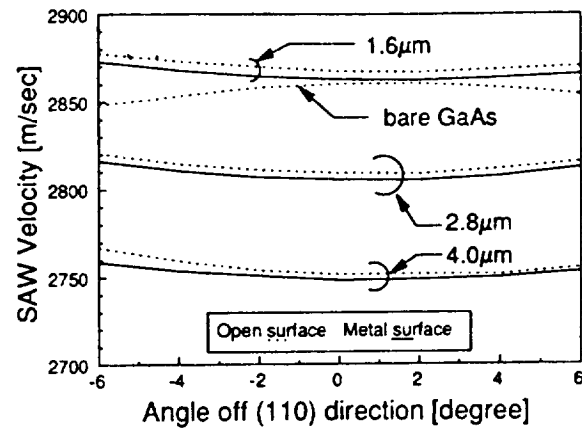
in detail in [23]. Due to the requirement of the water coupling between the SAM lens and the sample, the velocity measured is associated with leaky waves for which the waves leak their energy into the water. Thus, the measured velocity due to the water loading is slightly different from the one measured by the laser probe. The mass loading of the water is included in the $V(z)$ analysis. The LFBSAM can also measure both leaky SAW and pseudo SAW [23] which can exist in a $\{001\}$ -cut cubic crystal. The advantage of this method is the ability to measure the velocity profile in 360 degrees. However, the spatial resolution is limited (in this case an area of $300 \mu\text{m}$ by $1000 \mu\text{m}$) because of the defocussing requirements [23]. For the bare $\{001\}$ -cut GaAs, both leaky SAW and pseudo SAW profiles were reported [24].

Fig. 7(a) shows the measured velocity surfaces for the 0 (bare), 1.6, 2.8, and $4.0 \mu\text{m}$ ZnO film on $\{001\}$ -cut GaAs. It is seen that the ZnO films change the focusing behavior about the $\langle 110 \rangle$ direction into the nonfocusing one and the reverse inversions appear about the $\langle 100 \rangle$ direction. These results support our theoretical results and our laser probe measurements. Effects of the passivation layer ($0.1 \mu\text{m}$ SiO_2 or $0.2 \mu\text{m}$ Si_3N_4) for $1.6 \mu\text{m}$ ZnO and $\{001\}$ -cut GaAs on the velocity surface are provided in Fig. 7(b), and they also agree with the results obtained using the laser probe. Note that Fig. 7(a) and (b) are solely used to demonstrate the leaky SAW behavior about the $\langle 110 \rangle$ and $\langle 100 \rangle$ directions, the details of their behavior farther away from these angles are not shown.

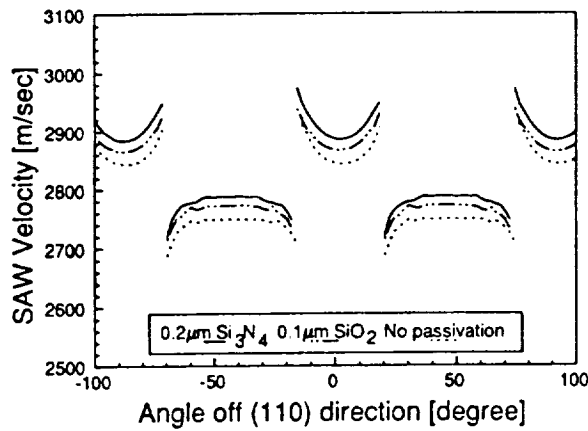
Experimental velocity surfaces with and without the metalization pad for small deviations about the $\langle 110 \rangle$ direction are given in Fig. 8. Because the velocity measurement accuracy performed by LFBSAM is better than $\pm 0.02\%$, in principle these data can be used to deduce the value of K^2 . Recently, the $V(z)$ analysis of LFBSAM measurements for a multilayered anisotropic structure has been reported [25]. However, due to a lack of understanding of the dielectric loading of the piezoelectric effect by the water, measurements of K^2 can not be carried out at present. On the other hand, the water is distilled and deionized to have a resistivity of $10^7 \Omega$. The conductivity effect of the water is mingled with its dielectric effect because the characteristic electromagnetic impedance



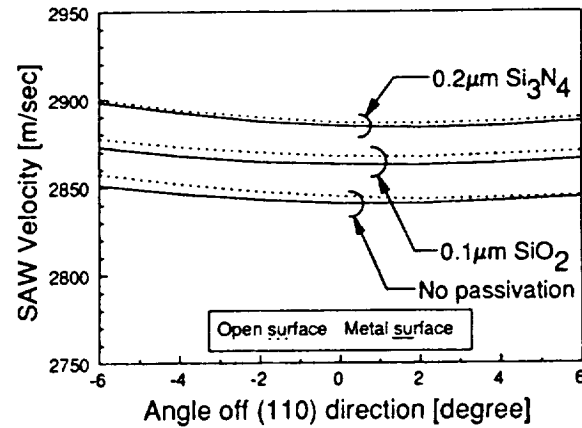
(a)



(a)



(b)



(b)

Fig. 7. Velocity surfaces measured by LFBSAM at 225 MHz. Numbers indicate the thicknesses of the ZnO film in μm . (a) Effect of ZnO film thickness with a $0.1 \mu\text{m}$ SiO_2 passivation layer, (b) Effect of passivation for $1.6 \mu\text{m}$ thick ZnO film.

Fig. 8. Velocity surfaces measured by LFBSAM for small deviations about (110) direction. Numbers indicate the thicknesses of the ZnO film in μm . (a) Effect of ZnO film thickness with a $0.1 \mu\text{m}$ SiO_2 passivation layer, (b) Effect of passivation for $1.6 \mu\text{m}$ thick ZnO film.

is equal to the root square of the permeability over the permittivity.

IV. REFLECTION PROPERTIES OF METALLIC GRATINGS

An exact theoretical analysis of the reflection properties is extremely difficult. The reflectivity and velocity shift of metal gratings on a single crystal can be analyzed by using Datta and Hunsinger's technique [3] if the metal thickness is small compared to the SAW wavelength. Considering that the metallic gratings effect the surface, Datta and Hunsinger's theory developed for a single material can be directly applied to a multilayered substrate by obtaining effective material constants of a single material equivalent to a multilayered substrate. The effective constants c_{eff} may be obtained by integrating multilayered material constants weighted with SAW energy distribution and normalized by the energy as follows.

$$c_{eff} = \frac{\int dy c(y) P(y)}{\int dy P(y)} \quad (4)$$

where $c(y)$ is a constant of the multilayered material tensor and $P(y)$ is the energy distribution as a function of the depth y . Since the substrate is piezoelectric, the mechanical, electro-mechanical, and electrical energy distribution are utilized for stiffness, piezoelectric, and permittivity tensor, respectively. The detailed expression for the energy distribution is available from the Laguerre polynomial technique [7] and is not reproduced here. The validity of using the effective constants may be checked by comparing between the SAW velocity of ZnO/GaAs and that of the equivalent single material as is shown in Fig. 9. The maximum difference between both velocities is 0.009.

The reflectivity of gratings can be divided into two components: the piezoelectric and the mechanical. The former depends on the piezoelectric coupling constant K^2 , and the latter depends on the thickness h of the metal in wavelength λ . The reflectivity r of a metal strip with a metallization ratio η at a frequency f can be given by

$$r = j \left[P_z(\eta) \frac{K^2}{2} + F_z \frac{h}{\lambda} \right] \sin \frac{\eta \pi f}{f_o} \quad (5)$$

where f_o is the center frequency, P_z is the piezoelectric scattering coefficient, and F_z is the first order mechanical

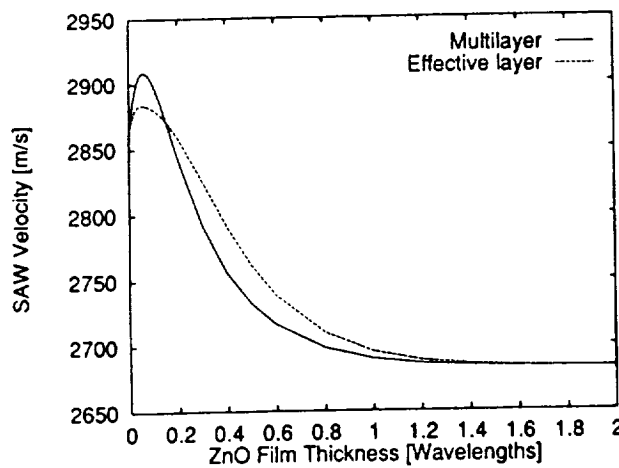


Fig. 9. A comparison between SAW velocity of multilayer (ZnO/GaAs) and that of effective single material.

scattering coefficient. P_z is the same for all substrate-electrode combinations and depends only on the metallization ratio. On the other hand, F_z depends only on the substrate and electrode materials and is independent of the metallization ratio. More detailed explanation about these parameters can be found in Hunsinger [3].

The value of P_z at f_o is given as a function of η by using the Legendre function [3]. It is to be -0.75 when the value of η is 0.5 . Fig. 10 shows the calculated values of K^2 and F_z for an aluminum grating with $\frac{1}{4}\lambda$ width and $\frac{1}{4}\lambda$ spacing (the same geometry for solid finger interdigital transducers (IDT's)) on ZnO/ $0.25\ \mu\text{m}$ Si_3N_4 /GaAs as a function of ZnO film thickness. As shown in Fig. 10, the value of K^2 increases from 0.07% (GaAs) to 1.0% (ZnO), and that of F_z decreases from 0.9 (GaAs) to 0.2 (ZnO) as the film thickness increases. Thus, it is possible for the piezoelectric scattering and the mechanical scattering to cancel each other due to their opposite signs. This zero reflectivity can not only reduce triple transit interference (TTI) of IDT's for some filter applications but also allows the use of a single finger IDT instead of a split finger IDT, thus increasing the frequency range of easily fabricated devices.

In order to verify this remarkable theoretical expectation, a set of characterization devices shown in Fig. 11 was designed to measure the effect of the gratings. The ZnO films were grown by dc sputtering method with thicknesses of 0.8 and $1.6\ \mu\text{m}$ with a $0.25\ \mu\text{m}$ PECVD grown Si_3N_4 passivation layer. The probe IDT's are 100λ wide where λ is $16\ \mu\text{m}$ with a center frequency around $180\ \text{MHz}$. Each of the probe IDT's is 20λ long, and split finger electrodes are employed in order to reduce internal reflections. Both shorted and open gratings with the width of 108λ are investigated, and each of the gratings consists of 200 strips with $\frac{1}{4}\lambda$ width and $\frac{1}{4}\lambda$ spacing. The distance from the center of the front two IDT electrodes with same polarity to the edge of the shorted grating is designed to be $(1/2n + 1/4)\lambda$ and $1/2n\lambda$, where n is an integer, to measure the sign of the reflectivity since the phase construction of a standing wave occurs at the distance of $(1/2n + 1/4)\lambda$ for a negative sign of reflectivity and $1/2n\lambda$ for a positive one [27].

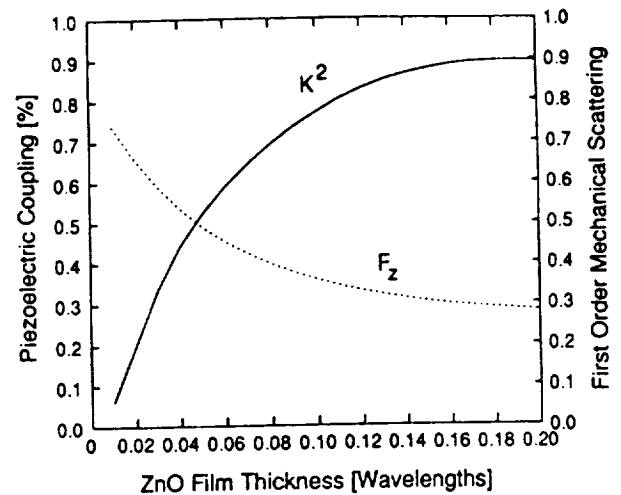


Fig. 10. Piezoelectric coupling K^2 and first order mechanical scattering coefficient F_z of aluminum gratings on ZnO/ $0.25\ \mu\text{m}$ Si_3N_4 /GaAs.

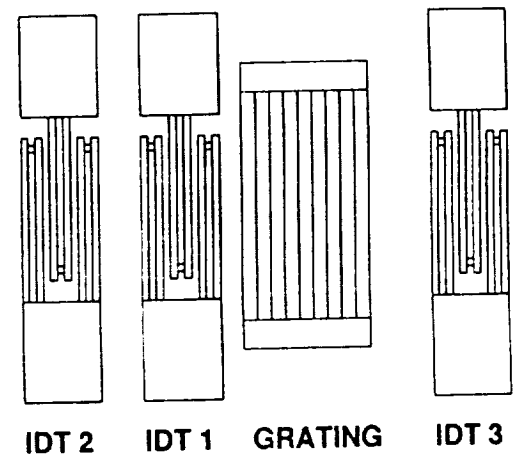


Fig. 11. Device configuration for reflectivity measurements.

The properties of open gratings were also investigated along with the shorted gratings. Unlike the shorted gratings, the SAW regeneration from the open grating should be considered in the analysis. At the center frequency, the reflectivity, r_o , of an open metal strip with the metallization ratio of 0.5 is given as [27]

$$r_o = r + j0.64K^2. \quad (6)$$

Due to this regeneration, the reflectivity of the open gratings always retains a positive value with a larger magnitude than that of the shorted gratings; thus, the open gratings may be suitable for reflectors or resonators, for which a strong reflectivity is desirable.

The identical IDT probes with free and metalized surface with 120λ wide and 200λ long metal pad instead of the gratings were also fabricated to calibrate the measured data. Moreover, from these calibration sets, the SAW properties of velocity, K^2 , and attenuation for the ZnO/GaAs substrates were also characterized. Three different thicknesses of the aluminum gratings were deposited with an e -beam evaporator. The measurements using a surface profile measurement apparatus showed they were 40 , 80 , and $120\ \text{nm}$, which

corresponded to 0.25, 0.50, and 0.75% of λ , respectively. The measurements were performed with an HP8753C network analyzer with a time-gating feature.

The velocities for the 0.8 μm and 1.6 μm ZnO thicknesses were 2928 m/sec and 2912 m/sec, respectively, from the center frequency measurements. The value of K^2 was also obtained by comparing the difference of the time delay from the IDT 1 and 2 to the IDT 1 and 3 between the free surface and the metalized surface. It was 0.73% and 1.1%, respectively, but their accuracies may be degraded by the inaccurate value for the free surface velocity obtained from the center frequency measurements due to the velocity shift under the IDT's. More accurate measurement using the laser probe or the LFBSAM was not attempted for these substrates, but the measurements for the other substrates are already discussed in the previous section. The propagation loss was also measured by comparing insertion loss of the IDT 1 and 2 pair and that of the IDT 1 and 3 pair with a similar manner described in Melloch and Wagers [18]. Note that the loss due to diffraction is not considered in this technique. The loss of the 0.8 μm and 1.6 μm thickness was obtained to be 0.7 dB/ μs and 0.8 dB/ μs on the metal pad, and 2.6 dB/ μs and 3.6 dB/ μs on the free surface, respectively. This difference between the free surface and the metal pad is due to the large diffraction of the ZnO/GaAs. The 120λ wide metal pad must act as a waveguide to confine the acoustic energy and reduce the insertion loss, resulting in the reduction of the propagation loss. Note that the results coincide well with the measurements by the laser probe, which are discussed in the previous section.

Fig. 12 shows typical unmatched frequency response between the IDT 1 and 2 pair for the two shorted gratings (as mentioned before, the difference of their distance from the IDT 1 is $\frac{1}{4}\lambda$) with the same metal thickness. Note that the phase construction and destruction are clearly shown about f_o . The reflectivity of the grating was obtained using the time gating technique described in Wright [28]. Comparing the response of the IDT 1 and 2 pair with the grating to that without gives a total reflectivity of the gratings assuming the distance between the IDT 1 and the grating is exactly correct for the total phase construction and destruction. Note that the propagation loss is negligible. The reflectivity of each electrode of the grating is given by [27]

$$|r| = \tanh^{-1}(R)/N \quad (7)$$

where R is the total reflectivity and N is the number of the electrodes.

Fig. 13(a) and (b) show the measured data and theoretical calculation for the reflectivity of an electrode in the shorted and the open grating at f_o as a function of the metal thickness normalized in λ . Prior work concerning gratings on GaAs can be found in the literature [29]. The value of K^2 was assumed to be 70% of the theoretical value based upon the measurements presented in the previous section. From the theoretical data, it is seen that for the shorted gratings the positive reflectivity on bare GaAs decreases, and eventually it becomes negative. The change in the reflectivity of the open gratings with film

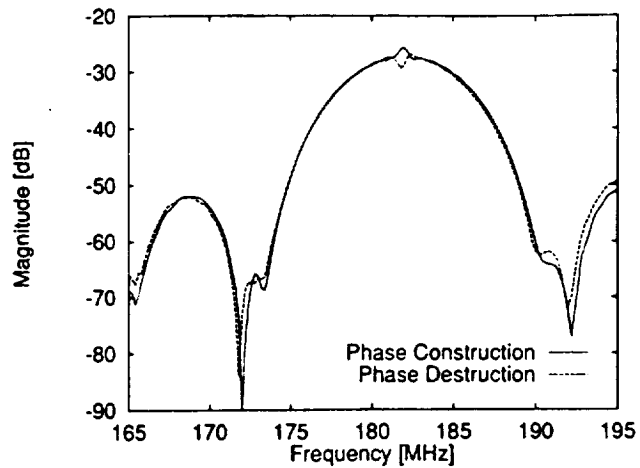


Fig. 12. Typical frequency response of two shorted gratings with the same metal thickness showing phase construction and destruction.

thickness is not so drastic as for the shorted gratings. On the other hand, the reflectivity on bare GaAs at f_o is given by $1.1h/\lambda$, and its value has a marginal difference between the shorted and the open gratings due to the very weakly piezoelectric regeneration. From the measurements, it should be noted that the sign of the reflectivity changes depending upon the metal thickness and zero reflectivity occurs at the metal thickness of 0.50% λ for 0.1 λ (1.6 μm) thickness of the ZnO film for the shorted grating while the reflectivity for the open grating is always positive. The measured data shows relatively good agreement with the theoretical prediction. The difference between the theory and the measurements may be due to the inaccurate effective material constants as well as the variation in the uniformity of the film and the metal thickness over the entire surface.

The measurement of velocity shift of the grating was attempted from the shift of the center frequency of the grating only response, f_{og} . The shift of f_{og} depending upon the metal thickness was not noticeable for both ZnO films, but the frequency shift between the shorted and the open grating was observed with the value of 0.05–0.25 MHz. It is known that the regeneration of SAW from the open grating causes an increase in the velocity by $0.25K^2$ at $\eta = 0.5$ [27]; thus, the measured values are comparable to the calculated ones of 0.18 MHz and 0.25 MHz for 0.8 μm and 1.6 μm thickness of the ZnO film, respectively.

V. SUMMARY

Theoretical and experimental data have been provided for the design of ZnO/GaAs based SAW devices, which may have the potential application for the monolithic integration of SAW and electronic devices. An analytical tool for SAW properties of a piezoelectric thin film was developed in this study, and its accuracy was verified with the experimental results of ZnO/GaAs substrates.

The SAW properties of both rf magnetron and dc triode sputtered ZnO films on {001}-cut (110)-propagating GaAs have been measured and compared to theoretical calculations. The passivation layer improves the ZnO film quality appreciably for both sputtering methods by reducing the propagation loss

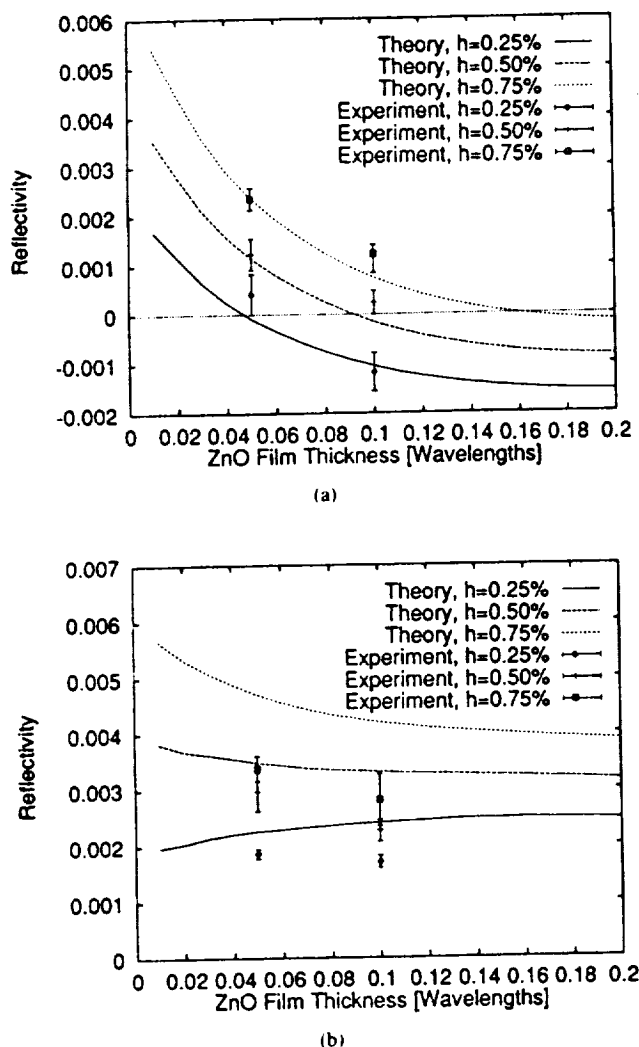


Fig. 13. Reflectivity of an aluminum electrode with $\frac{1}{4}\lambda$ width and spacing $\frac{1}{4}\lambda$ at f_0 , where the metal thickness in λ is represented by h : (a) shorted grating, (b) open grating.

0.5–1.3 dB/ μ s at 240 MHz depending upon the ZnO film thickness. The best quality of film observed has been 1.6 μ m dc triode sputtered ZnO film with 0.2 μ m Si_3N_4 , for which the propagation loss was 0.9–1.0 dB/ μ s at 240–280 MHz. Its value of K^2 would be larger than that of the film without the passivation layer, 0.65–0.75%. Compared with bare GaAs, this particular film substrate has approximately 0.6% lower SAW velocity, at least 10 times larger K^2 , and 0.6 dB/ μ s more propagation loss.

The velocity surfaces of the ZnO/GaAs substrates have been measured using a knife-edge laser probe and a LFBSAM and have been theoretically predicted. It is found that for certain thicknesses of ZnO film the velocity surfaces about the (110) direction are concave upward in contrast with the concave downward curvature of the bare GaAs. This causes the SAW to diffract as soon as it leaves the transducer. It is also found theoretically that a c -axis oriented hexagonal film such as AlN and ZnO can be utilized to control the diffraction of a cubic substrate such as GaAs, GaP, InAs, and InP by varying the film thickness.

The reflectivity of shorted and open metal gratings has been analyzed and measured. The analysis showed that zero reflectivity could be obtained by a combination of the ZnO film and the metal thickness and the metallization ratio. Experimentally, zero reflectivity at the center frequency was observed at 0.5% λ thickness of a shorted aluminum grating with $\frac{1}{4}\lambda$ width and $\frac{1}{4}\lambda$ spacing on 1.6 μ m thick ZnO film with 0.25 μ m thick Si_3N_4 passivation layer over GaAs.

ACKNOWLEDGMENT

The authors wish to thank Prof. S. R. Stock of Georgia Tech for the X-ray diffraction measurement.

REFERENCES

- [1] F. S. Hickernell, "ZnO processing for bulk- and surface-wave devices," in *Proc. IEEE Ultrason. Symp.*, 1980, pp. 785–794.
- [2] F. S. Hickernell, "Piezoelectric thin-film materials," in *Proc. Int. Symp. SAW Devices for Mobile Comm.*, 1992, pp. 31–38.
- [3] B. J. Hunsinger, "Research to provide a theoretical determination of surface acoustic wave velocity and impedance differences between metal strips and free surface regions of metallic gratings," Technical Report, Rome Air Development Center, 1981.
- [4] G. R. Adams, J. D. Jackson, and J. S. Heeks, "Monolithic ZnO-GaAs acousto-electric devices," in *Proc. IEEE Ultrason. Symp.*, 1980, pp. 109–112.
- [5] T. S. Hickernell, F. M. Fliegel, and F. S. Hickernell, "The elastic properties of thin-film silicon nitride," in *Proc. IEEE Ultrason. Symp.*, 1990, pp. 445–448.
- [6] T. S. Hickernell and F. S. Hickernell, "The elastic properties of PECVD silicon oxynitride films on gallium arsenide," in *Proc. IEEE Ultrason. Symp.*, 1991, pp. 51–54.
- [7] Y. Kim and W. D. Hunt, "Acoustic fields and velocities for surface-acoustic-wave propagation in multilayered structures: An extension of the Laguerre polynomial approach," *J. Appl. Phys.*, vol. 68, pp. 4993–4997, 1990.
- [8] B. Hadimioglu, S. A. Elrod, D. L. Steinmetz, M. Lim, J. C. Zesch, B. T. Khuri-Yakub, E. G. Rawson, and C. F. Quate, "Acoustic ink printing" in *Proc. IEEE Ultrason. Symp.*, 1992, pp. 929–935.
- [9] W. D. Hunt, Y. Kim, and F. M. Fliegel, "A synopsis of surface acoustic wave propagation on {100}-cut (110)-propagating gallium arsenide," *J. Appl. Phys.*, vol. 69, pp. 1936–1941, 1991.
- [10] J. Temmyo and S. Yoshikawa, "Aluminum evaporation conditions for SAW interdigital transducers," *IEEE Trans. Son. and Ultrason.*, vol. 27, pp. 219–221, 1980.
- [11] V. E. Steel, W. D. Hunt, M. A. Emanuel, J. J. Coleman, and B. J. Hunsinger, "Surface acoustic wave properties of aluminum arsenide," *J. Appl. Phys.*, vol. 66, pp. 90–96, 1989.
- [12] W. D. Hunt, R. L. Miller, and B. J. Hunsinger, "Slowness surface measurements for zero- and five-degree [100]-cut GaAs," *J. Appl. Phys.*, vol. 60, pp. 3532–3538, 1986.
- [13] A. J. Slobodnik, E. D. Conway, and R. T. Delmonico, *Microwave Acoustics Handbook*. Air Force Cambridge Research Laboratory, 1973, vol. 1A.
- [14] V. M. Bright, Y. Kim, and W. D. Hunt, "Study of surface acoustic waves on the {110} plane of gallium arsenide," *J. Appl. Phys.*, vol. 71, pp. 597–605, 1992.
- [15] Y. Kim, W. D. Hunt, Y. Liu, and C. K. Jen, "Velocity surface measurements for ZnO films over {001}-cut GaAs," *J. Appl. Phys.*, vol. 76, pp. 1455–1461, Aug. 1, 1994.
- [16] A. J. Slobodnik, "Materials and their influence on performance," in *Acoustic Surface Waves, Topics in Applied Physics*, A. A. Oliner, Ed. New York: Springer-Verlag, 1978, vol. 24, chapter 6.
- [17] F. S. Hickernell, "Surface acoustic wave propagation loss in zinc oxide films," in *Proc. IEEE Ultrason. Symp.*, 1982, pp. 325–328.
- [18] M. R. Melloch and R. S. Wagers, "Propagation loss of the acoustic pseudosurface wave on (ZX)45° GaAs," *Appl. Phys. Lett.*, vol. 43, pp. 1008–1009, 1983.
- [19] T. W. Grudkowski and M. Gilden, "Realization of temperature-compensated GaAs surface acoustic wave delay lines," *Appl. Phys. Lett.*, vol. 38, pp. 412–413, 1981.
- [20] W. D. Hunt and B. J. Hunsinger, "A precise angular spectrum of plane waves diffraction theory for leaky wave materials," *J. Appl. Phys.*, vol. 64, pp. 1027–1032, 1988.

- [21] M. G. Cohen, "Optical study of ultrasonic diffraction and focusing in anisotropic media," *J. Appl. Phys.*, vol. 38, pp. 3821-3828, 1967.
- [22] B. A. Auld, *Acoustic Fields and Waves in Solids*. New York: Wiley, 1973, vol. 1.
- [23] J. Kushibiki and N. Chubachi, "Material characterization by line-focus-beam acoustic microscope," *IEEE Trans. Ultrason. Ferroelec. Freq. Cont.*, vol. 32, pp. 189-212, 1985.
- [24] A. Briggs, *Acoustic Microscopy*. New York: Oxford University Press, 1992, chap. 11.
- [25] Y. C. Lee, J. O. Kim, and J. D. Achenbach, "V(z) curves of layered anisotropic materials for the line-focus acoustic microscope," *J. Acoust. Soc. Amer.*, vol. 94, pp. 923-930, 1993.
- [26] P. S. Cross, "Reflective arrays for SAW resonators," in *Proc. IEEE Ultrason. Symp.*, 1975, pp. 241-244.
- [27] S. Datta, *Surface Acoustic Wave Devices*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
- [28] P. V. Wright, "Modeling and experimental measurements of the reflection properties of SAW metallic gratings," in *Proc. IEEE Ultrason. Symp.*, 1984, pp. 54-63.
- [29] T. W. Grudkowski, G. K. Montress, M. Gilden, and J. F. Black, "Integrated circuit compatible surface acoustic wave devices on gallium arsenide," *IEEE Trans. Microwave Theory and Technique*, vol. 29, pp. 1348-1356, 1981.



Yoonkee Kim received the B.S. and the M.S. degrees in electrical engineering from Seoul National University in 1981 and Korea Advanced Institute of Science and Technology in 1984, respectively. He received the M.S. and Ph.D. degrees in electrical engineering from Georgia Institute of Technology in 1991 and 1993, respectively.

From 1984 to 1986 he was employed as a Research Engineer in the Central Research Laboratory of the Goldstar Inc., in Seoul. During his graduate studies at Georgia Institute of Technology, his research interest was in SAW and BAW devices as well as ACT devices. From 1993, he worked as a Postdoctoral Fellow in Georgia Institute of Technology. He has several publications in the microacoustics area.



William D. Hunt received the B.S. degree in 1976 from the University of Alabama, the S.M. degree from M.I.T. in 1980, and the Ph.D. degree in 1987 from the University of Illinois, Champaign-Urbana, all in electrical engineering.

He worked for Harris Corporation from 1976 to 1978, and for Bolt Beranek and Newman from 1980 to 1984, and joined the faculty of Georgia Tech in 1987 where he is currently an Associate Professor. His research interests include SAW and ACT devices as well as transducers for biomedical

ultrasound.

He was a Rhodes Scholar Finalist in 1975, received a DuPont Young Faculty Award in 1988 and was named an NSF Presidential Young Investigator in 1989. In addition, he has been selected for *Who's Who in the South and Southwest* and *Who's Who in American Education* and was named as one of the 1994 Distinguished Engineering Fellows by the University of Alabama College of Engineering.

Frederick S. Hickernell (M'64-SM'70-F'91), for photograph and biography, please see p. 330 of this issue.



Robert J. Higgins graduated from Mississippi State University in 1976 with a B.S.E.E. degree. In 1981 he received his M.S.E.E. degree from Florida Atlantic University.

In 1976 he joined Motorola where he designed transmission line filters for portable communications systems. He is presently a Member of the Technical Staff in Motorola's Radio Products Group Applied Research Department.

Cheng-Kuei Jen, photograph and biography not available at time of publication.

5.3 An Ultra-low power ACT Device:

In this section we describe the work done to develop an ACT device which would not have an IDT at all but would instead have an acoustic ring oscillator comprised of two banks of slanted reflectors, similar to the arrays for Reflective Array Compressors, and an acoustoelectric amplifier and an ACT device. This approach would allow for a single channel ACT device to operate at a power consumption of about 10mW with a 2Volt bias. If there are efforts in the future to revive ACT technology, this would be the best approach as it overcomes virtually all of the technology's shortcomings which have kept it out of commercial systems. This work is described in the attached paper:

Cameron, T.P. and Hunt, W.D., "Slanted Reflection Gratings on Gallium Arsenide," *Journal of Applied Physics*, vol. 84, no. 4, pp. 2214-2218, August 1998.

Slanted Reflection Gratings on Gallium Arsenide

by

Thomas P. Cameron and William D. Hunt

School of Electrical and Computer Engineering, and The Microelectronics Research Center
Georgia Institute of Technology, Atlanta, Georgia 30332

Abstract

In this paper we present the first experimental study of the behavior of a slanted reflector array on {100}-cut, <110>-propagating GaAs. The thorough examination of both Al stripes and grooves on GaAs unveiled previously unknown information about the behavior of slanted gratings on {100}-cut GaAs. We discovered that etched grooves provide a strong reflection coefficient ($C=0.29$) while Al stripes do not ($C=0.014$). These data measurements of the reflection characteristics for slanted gratings on GaAs. Potential device applications include an ultra-low power ACT device which does not require an IDT.

I. INTRODUCTION

Arrays of acoustic reflectors on surface acoustic wave (SAW) devices have been well studied and many applications have been derived from the use of both normal and oblique incidence reflectors, including a whole range of low-loss SAW filters and resonators, and reflective array compressors (RACs). The low-loss filters and resonators have been primarily based on normal incidence reflections. The RAC devices have primarily been based on oblique incidence reflections from a pair of slanted mirror-image reflector banks and it is this type of reflector array which is the subject of this paper. Although 90° reflectors have been well studied for the more common piezoelectric substrates such as quartz and LiNbO_3 [1], to the best of the authors' knowledge there exists no previous experimental study on the behavior of slanted reflectors on GaAs. The primary technological motivation for this work was to develop an efficient reflector which could be used as a key component in a acoustic ring oscillator (ARO) for an Acoustic Charge Transport (ACT) device [2]. It is possible that with a pair of reflector banks and an acoustoelectric amplifier section, an ARO can be achieved such that sufficient potential for charge transport can be achieved with only a few volts DC applied to the acoustoelectric amplifier section. For this ARO, there would be no need for an IDT which has consumed so much RF power and precious real estate in previous ACT device architectures. The reflector

banks not only circulate the energy around the loop, but also define the operating frequency of the oscillator and the bandwidth of the loop. In order to design an efficient loop, it was first necessary to understand the operation of slanted 90° reflectors on GaAs.

It has been demonstrated experimentally in this paper, that the generation of bulk modes using grooves on GaAs appears to be nonexistent, and efficient, low loss reflectors on GaAs can be readily achieved.

We begin with the theory of slanted reflectors. Approximations are made which enable the reflection coefficients for single steps to be experimentally determined. The test structures used in this study are then briefly described and the experiments to determine the reflection coefficients are described. As a result, the reflection coefficients for a single step discontinuity for an Al stripe and a groove are experimentally determined.

II. THEORY

The operation of slanted reflector gratings is illustrated in Figure 1. The SAW is launched by the IDT at the top left. The SAW is partially reflected by each stripe through 90° to the bottom track where the same process takes place to direct the SAW to the IDT at the bottom left. The predominant attributes of the array are the stripe period, p_z , and the stripe width, a_z , and the height of the step, h . These values not only set the synchronous frequency of the array, but also the efficiency of the reflection process.

The angle, ϕ , of the slanted reflectors must be set individually for a given cut of a material to reflect the waves at 90 degrees. For anisotropic substrates ϕ is generally not 45° but because GaAs is a cubic material the velocity for the incident and reflected waves in the grating are equal. Further the temperature characteristics for the wave are invariant under 90° rotations in the {001}-plane, hence the appropriate value for ϕ is 45° and the optimum angle should not change considerably with temperature.

For the reflection of a wave of amplitude A on a single groove, the reflection coefficients for an up step

Γ_u , a down step Γ_d and the transmission coefficient τ , for either step are from [1]

$$\Gamma_u = r - j\frac{\hat{B}}{2} \quad (1)$$

$$\Gamma_d = -r - j\frac{\hat{B}}{2} \quad (2)$$

$$\tau = 1 - \frac{r^2}{2} - j\frac{\hat{B}}{2} \quad (3)$$

where r has been defined as the impedance discontinuity, which is proportional to the step height, i.e. $r = C\frac{h}{\lambda}$, \hat{B} is the energy storage term proportional

to the square of the step height, i.e. $\frac{\hat{B}}{2} = C'\left(\frac{h}{\lambda}\right)^2$ where λ is the acoustic wavelength, and C and C' are the first and second order coefficients of mechanical scattering which we will determined experimentally.

The reflection coefficient for a groove, Γ_g , can be derived by following the path of the wave and using equations (1), (2) and (3). When the terms of order $(h/\lambda)^2$ are neglected, then Γ_g was expressed as [1]

$$\Gamma_g = -2je^{-j\theta}(r \sin \theta + \frac{\hat{B}}{2} \cos \theta) \quad (4)$$

where

$$\theta = \frac{ka_z}{2} + \frac{\hat{B}}{2} \quad (5)$$

and k is the wavenumber in the z direction. For the case for our experiments $a_z = \lambda/2$, so θ becomes:

$$\theta = \frac{\pi}{2} + \frac{\hat{B}}{2} \quad (6)$$

Now, for the metal thickness under consideration ($h/\lambda < 0.04$), it can be assumed that $\frac{\hat{B}}{2} \ll \frac{\pi}{2}$, so θ is taken as $\pi/2$, making the magnitude of Γ_g ,

$$|\Gamma_g| = 2r \quad (7)$$

In the case where $a_z = \lambda$, then $\theta = \pi$, which leads to

$$|\Gamma_g| = \hat{B} \quad (8)$$

Equations (7) and (8) allow approximate values of C and C' to be determined experimentally by using gratings with widths of $\lambda/2$ and λ respectively. An experiment, conducted for grooves on Y-Z LiNbO₃ by Melngailis and Li [4], showed a significant difference in constants for normal and oblique incidence. Hence, the values predicted for normal

incidence cannot generally be applied to the oblique case, which supports the findings of this paper.

III. EXPERIMENTAL RESULTS

A. Test Structure

Because the maximum reflection from an oblique grating is attained for $a_z = \lambda/2$ [5], we set out to determine the value of the impedance mismatch, r , for 45 degree grooves and Al stripes on GaAs. We designed a number of devices illustrated schematically in Figure 1. The test structure consisted of three IDTs and two reflector banks in between. The SAW was launched by the IDT at the upper left. The transmission through the reflector (T31) was measured by the IDT at the right, and the reflection from the pair of reflector banks (T21) was measured by the IDT at the bottom left. An alternative technique would have been to have only one reflector bank and a large IDT rotated 90 degrees on the bottom. However, the structure we utilized more closely matches the reflector structure to be used in a ring oscillator, and hence provided more useful information in this case. Due to reciprocity, it was assumed that the responses of the gratings were identical, allowing the reflection coefficient of a single reflector bank to be deduced from the measurement.

Ten different test structures were included on mask set AE3. The IDTs on these devices were split finger IDTs with a finger width of 1.5 μm , designed to operate at about 238.5 MHz. The split finger IDT was chosen because it produces no appreciable internal reflections which could interfere with the measurement of the grating characteristics. The reflector banks all had a stripe and gap width of 6 μm in the direction of propagation, corresponding to $a_z = \lambda/2$.

All stripes were inclined 45°. There were two sets of 5 reflectors with lengths varying from 200 to 1000 stripes, one set with open stripes and one set with all the stripes shorted. The IDTs and reflectors were laid out on two different plates, so that the metal thickness of the two structures could be different, and so that grooves could be investigated using the same mask set.

B. Reflection Coefficient for Al Stripes

Several sets of devices were fabricated for this experiment on 2 inch semi-insulating (SI) {001}-cut, <110>-propagating GaAs wafers. The first experiment performed was on Al stripes with a 100 Å Cr adhesion layer. Stripes were investigated first as it was hoped

that a ring oscillator device could be simplified by having a single metallization layer for both the IDTs and the reflectors. Three Al thicknesses were deposited using a thermal evaporator. The metal was subsequently measured using the DECTAK 3030 auto profiling system to the nearest 100 Å. In this experiment the thickness of the Cr was neglected as the ratio of Al to Cr is large for all three cases.

The IDT to IDT response was a measurement made between two IDTs with no reflectors between them as a benchmark on each wafer. The other two measurements made were T21 and T31, to evaluate the reflection and transmission coefficients of the grating. The reflection coefficient of a single grating, which was of most interest here, was calculated from the log magnitude of the S21 measurement of T21

$$|\Gamma_{\text{grat}}| = \sqrt{10^{\frac{|T21|}{20}}} \quad (9)$$

We assumed each grating to be identical and hence would contribute equally to the net reflection coefficient.

From the transmission matrix model of Datta [6], the reflection coefficient of a grating (Γ_{grat}), whether slanted or normal, was approximated from the impedance mismatch $\Delta z/z$ simply as

$$|\Gamma_{\text{grat}}| = \tanh(N_g \frac{\Delta z}{z}) \quad (10)$$

where, to equate terminology, r used in this paper is equivalent to Datta's $\Delta z/z$ if piezoelectric shorting is ignored. Since the piezoelectric coupling for this cut of GaAs is so small, this is a reasonable assumption.

From Equation (10), it was then straightforward to compute the reflection coefficient of a single step, r , from the measured data as

$$|r| = \frac{1}{2N_{\text{grat}}} \tanh^{-1} |\Gamma_{\text{grat}}| \quad (11)$$

Calculated values $|\Gamma_{\text{grat}}|$ for each of the devices in the experiment have been compiled in Table 1 from the measurement of $|T21|$. There are some interesting trends which are observed from the data in this table. First of all, the value of $|\Gamma_{\text{grat}}|$ generally increases with metal thickness as expected. No conclusive difference was observed between open and shorted stripes as could be expected because of the low value of K^2 . The most interesting trend is the decrease in $|\Gamma_{\text{grat}}|$ for the long gratings, particularly for the thickest metal. We

attribute this phenomenon, which is contrary to the expected behavior, to secondary reflections in the grating.

C. Multiple Reflections

By examining the paths of triply-reflected waves, it was observed that at the resonant frequency of the grating these waves are 180° out of phase with respect to the singly-reflected waves [3]. In Figure 2, which has been borrowed from [4], the effect of triply-reflected waves on the time response of the T21 has been sketched. For the case of only singly-reflected waves, the time response would be a simple rectangular pulse train. The triply-reflected waves produce a triangular shaped pulse train. The net time response, then, is the convolution of the two pulse trains, resulting in a non-rectangular pulse train twice as long as would be expected if there were no multiple reflections. The principal effects of the multiple reflections are a narrowing of bandwidth of the frequency response and a decrease in the magnitude of the reflection coefficient of the grating.

The multiple reflections model, explains precisely what was observed in the experimental data for the slanted Al gratings. For example, the time domain response for a 600 stripe grating with an Al thickness of 3800 Å has showed an impulse response of T21 that followed the behavior just described. This leads two insights into the design of Al gratings. Longer gratings are not necessarily more efficient and thicker metal does not necessarily produce a higher net reflection coefficient for the grating.

D. First Order Mechanical Scattering for Al Stripes

Given the multiple reflections in the longer IDT, only the 200 stripe reflector bank was used in this study to approximate a value of C for Al stripes. Using equation (11), the value of $|r|$ was calculated for the three metal thicknesses for a reflector bank with 200 open stripes. These data are plotted in Figure 3 along with a linear approximation which was fitted to the data to approximate the slope of the line. The linear approximation was described by $0.0009 + 0.014 h/\lambda$. It was not possible to match a linear curve to the data without the constant term. From the work of Datta [6], this constant has been attributed to the reflection due to piezoelectric shorting. This value was determined by Datta for normal incidence gratings, but no previous

value has been reported for oblique incidence gratings. For normal incidence gratings, this value has been calculated as $-0.7 \cdot K^2/2$, which for this cut of GaAs would be 0.00049. This is on the same order of magnitude as the constant experimentally determined by the authors, but no more specific explanation of the difference has been attempted here.

We computed the value of C for shorted Al stripes to be approximately $C = 0.014$, which is a low value compared to that for normal incidence (0.9) predicted by Datta. Using this value of $C=0.014$, in Equation (11) we predicted that a metal thickness of $4.6 \mu\text{m}$ would be necessary to obtain $|\Gamma_{\text{gr}}| = 0.99$ for $N_{\text{gr}} = 200$. A value of only $0.34 \mu\text{m}$ was calculated for $N_{\text{gr}} = 1000$, but from the experimental data this was obviously not the case, due to the high level of loss caused by the multiple reflections in the grating.

From this exercise, we concluded that the desired efficiency for a ring oscillator could not be achieved using a grating with Al stripes. No other metal was considered for further investigation but it is possible that a different metal such as Au or a combination of metals as demonstrated by Hunt and Hunsinger for normal incidence gratings [8], might provide better performance. To summarize, the reflection coefficient from single step in an Al grating has been found to be small, which leads to the need for either excessively thick metal layers or very long reflector banks. The necessary Al thickness is not practical for SAW device processing and the high loss introduced by multiple reflections in long gratings makes this an undesirable approach. This lead us to our investigation of grooves.

E. First Order Mechanical Scattering Coefficient for Grooves

We conducted experiments on grooves similar to those previously described for stripes. The wafers were processed in the same manner as the wafers used for stripes up to the point where the stripes were patterned with the photoresist. At this point the wafers grooves were etched using a $\text{H}_3\text{PO}_4/\text{H}_2\text{O}_2/\text{H}_2\text{O}$ solution at a ratio of 3:1:50. The etch was performed at room temperature ($\sim 22^\circ\text{C}$). The etch rate was measured to be approximately 800 \AA per minute and the only peculiarity of the etch was that patterns with the shorting bar along the edge were reproducible, whereas the open strip patterns etched sporadically and could

not be relied on to give a uniformly etched grating. Given the small dimensions of the grooves, the thickness could only be measured in the large busbar region. It was observed using a scanning electron microscope, that the etch was of uniform depth from the large busbar region into the smaller grooves so that this measurement gave a valid representation of the groove depth. SEM pictures of the grooves showed that the profile of the sidewall was inclined approximately 45 degrees. It may be possible that a different etch could have produced straight sidewalls, but this profile produced a strong reflection coefficient, so no further investigation was performed.

The same measurements and data analysis were performed on the grooves as previously described for the Al stripes. Five samples were fabricated with groove depths ranging from 1000 \AA to 4300 \AA . The calculated values of $|\Gamma_{\text{gr}}|$ have been tabulated in Table 2. It was immediately obvious from the data that grooves provided a much higher reflection coefficient than did the Al stripes. It was initially concluded, that stripes were less susceptible to the effects of multiple reflections than were the stripes. It was observed that the reflection coefficient did not drop off as dramatically for the long grating arrays. However, further analysis contradicted this conclusion. Upon further observation it became clear that because of the high reflectivity of the grooves, the wave penetration into the grating length was much less than in the Al stripe arrays. Hence, the effects of multiple reflections are not as apparent as in the Al stripe arrays.

The wave penetration was further investigated using a knife-edge laser probe to establish the effective length of the gratings so that the value of r could be deduced from the data. The knife-edge laser probe system is capable of measuring the relative magnitude and phase of the SAW as a function of position in the device under test. For the Al stripes, this was not an issue, as the reflection coefficient was weak. For the deeper grooves, it was necessary to establish an effective length, N_{gr} in order to approximate the value of $|r|$, as it was found that the SAW did not penetrate the full 200 grooves. The method used to scan along the direction of propagation with the laser spot centered transversely in the grating. The magnitude of the SAW was measured at several points to establish the

penetration depth. The measured data for a 200 groove grating with 4300 Å groove depth has been plotted in Figure 4. It was observed from this measurement that the SAW amplitude dropped off quickly, indicating a strong reflection coefficient. The effective grating length was taken to be the point where the amplitude dropped to one tenth of the incident SAW amplitude, which in this case was estimated to be $N_{gr} = 70$.

Using this data, combined with the measured value of $|\Gamma_{gr}|$, the value of $|r|$ was approximated as before. The experimentally calculated values of $|r|$ have been plotted against normalized groove depth in Figure 5 along with a linear curve fit. In this case, the slope, C , of the curve was approximated at 0.29 and no constant was necessary to obtain a good fit. This supports the earlier hypothesis that the constant was needed to account piezoelectric shorting, which does not exist for grooves.

IV. SUMMARY AND CONCLUSIONS

The thorough examination of both Al stripes and grooves on GaAs unveiled previously unknown information about the behavior of slanted gratings on {100}-cut GaAs. Although it was initially hoped that the Al stripes would have made efficient gratings, this did not turn out to be the case. It was, however, discovered that etched grooves provided a strong reflection coefficient. Previously unreported values for mechanical reflection coefficients have been approximated in this paper from measured data. Although the experiments were affected somewhat by multiple reflections, these constants provide a good first order design rule for the design of slanted gratings on GaAs. Based on this study it seems feasible that an acoustic ring oscillator can be fabricated on GaAs using groove reflector arrays.

REFERENCES

- [1] J. Melngailis, R.C. Williamson, J. Holtham and R.C. Li, "Design of Reflective-Array Surface Wave Devices", *Wave Elec.*, vol 2, 1976, pp 177-198.
- [2] T.P. Cameron, *Low-Voltage SAW Amplifiers on Multilayer GaAs/ZnO Substrates*, Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA 1996.
- [3] D. Morgan, *Surface Acoustic Wave Devices for Signal Processing*, Elsevier, Oxford, 1985.
- [4] J. Melngailis and R.C. Li, "Measurement of Impedance Mismatch and Stored Energy for

Right-Angle reflection of Rayleigh Waves from Grooves on Y Cut LiNbO_3 ", *Proc. 1975 IEEE Ultrason. Symp.*, pp 426-429.

[5] K. Yamanouchi and F. Kadosawa, "An Electroacoustic Surface-Wave Convolver Fabricated From a Thin Piezoelectric Film and a Semiconductor", *Elec. Comm. Jap.*, Part 2, vol.75, no. 12, 1992.

[6] S. Datta, *Surface Acoustic Wave Devices*, Prentice-Hall, Englewood Cliffs, N.J., 1986, p.184.

[7] F. Hickernell, "Piezoelectric Thin-Film Materials", *Proc. Int Symp. SAW Devices for Mobile Comm.*, 1992, pp 31-38.

[8] Hunt, W. D. and Hunsinger, B. J., "Reduction of metallic grating impedance and velocity perturbations via buried bimetallic electrode geometries," *Journal of Applied Physics*, vol. 63, no. 1, pp. 5237-5245, June 1988.

Table 1: Values of $|\Gamma_{gr}|$ calculated for Al stripes from measured data.

	Thickness (Å)		
N_{gr}	1,000	1,500	5,800
200Open	0.42	0.47	0.56
200Short	0.43	0.46	0.54
400O	0.5	0.53	0.65
400S	0.48	0.52	0.62
600O	0.56	0.56	0.70
600S	0.54	0.54	0.71
800O	0.56	0.57	0.63
800S	0.55	0.54	0.65
1000O	0.57	0.42	0.51
1000S	0.52	0.41	0.50

Table 2: Values of $|\Gamma_{gr}|$ calculated for grooves from measured data

	Groove Depth (Å)		
N_{gr}	1,000	1,500	2,700
200	0.65	0.71	0.86
400	0.72	0.78	0.97
600	0.72	0.83	0.96
800	0.75	0.87	0.95
1000	0.76	0.80	0.95

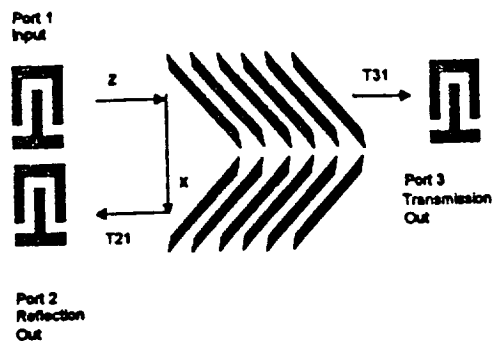


Fig. 1: Slanted reflective array test structure

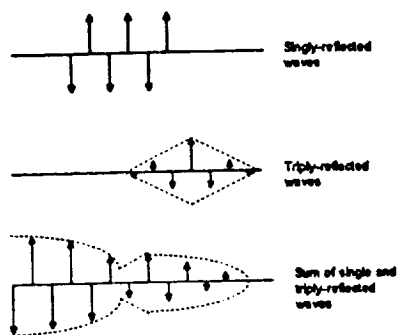


Fig. 2: Effects of multiple reflections on $T_{21}(t)$ [4]

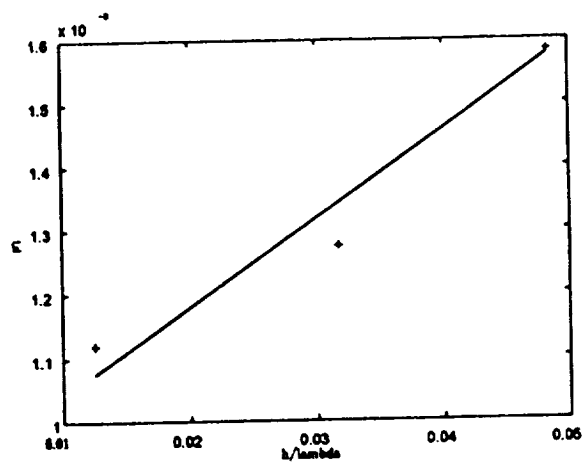


Fig 3: $|r|$ measured for 200 Al stripes h/λ (+) and curve fit used to calculate C (-)

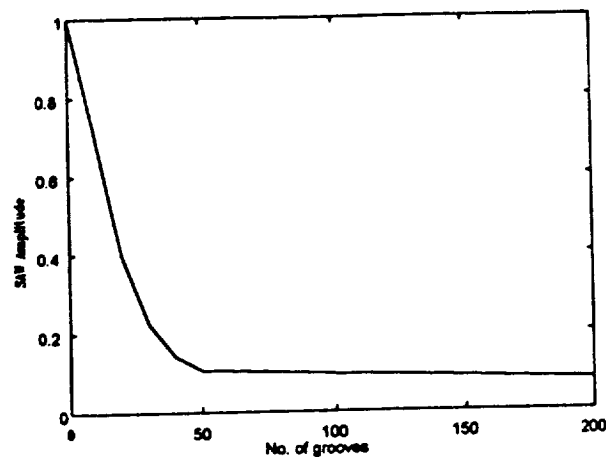


Fig. 4: Normalized SAW amplitude versus depth into the grating as measured using the laser probe.

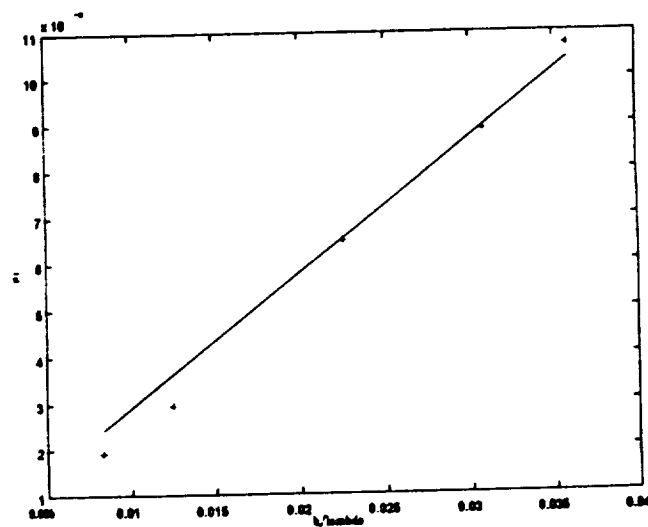


Fig. 5: Reflection coefficient of a single step, $|r|$, calculated from measured data for 200 grooves versus h/λ (+) and the fit curve used to calculate C (solid).

5.4 SAW Mode Profile Design and Analysis:

Given architecture of the ACT Imager, the photodiode array presents itself acoustically as a multichannel device with a plethora of waveguides. In an effort to develop the tools for this complex structure, we introduced stack matrix theory as a way of predicting the mode profiles for arbitrary waveguide structures. This approach has been adapted as the industry standard for the design of SAW Waveguide-coupled resonators. Once the SAW velocities in the various materials of the ACT imager structure are known, the stack matrix theory can be used to predict mode profiles and to design the structures' acoustic arrangement to optimize the imager structure acoustic operation. This work is described in the attached paper:

Hunt, W.D., Cameron, T., Saw, J.C.B, Kim, Y. and Suthers, M., "Mode profiles in waveguide-coupled resonators," *Journal of Applied Physics*, vol. 74, no. 8, pp. 4886-4893, October 1993.

Mode profiles in waveguide-coupled resonators

William D. Hunt, Tom Cameron,^{a)} John C. B. Saw,^{a)} Yoonkee Kim,
and Mark S. Suthers^{a)}

*School of Electrical Engineering, The Microelectronics Research Center, Georgia Institute of Technology,
Atlanta, Georgia 30332*

(Received 7 October 1992; accepted for publication 16 June 1993)

Surface acoustic wave (SAW) waveguide-coupled resonators are of considerable interest for narrow-band filter applications, though to date there has been very little published on the acoustic details of their operation. As in any resonator, one must fully understand its mode structure and herein we study the SAW mode profiles in these devices. Transverse mode profiles in the resonant cavity of the device were measured at various frequencies of interest using a knife-edge laser probe. In addition we predict the mode profiles for the device structure by two independent methods. One is a stack-matrix approach adapted from integrated optics and the other is a conventional analytical eigenmode analysis of the Helmholtz equation. Both modeling techniques are in good agreement with the measured results.

I. INTRODUCTION

Surface acoustic wave (SAW) waveguide-coupled (WGC) resonators, as shown in Fig. 1, are a class of devices which are proving to be quite useful in narrow-band filtering applications. Though these devices were proposed quite some time ago¹ there have since been only a few papers published on these devices. Tanaka *et al.*² have published the most comprehensive analysis of the WGC resonator to date, using an analytical modal analysis to predict the resonant frequencies of the device and the frequencies of the spurious modes. Their work provides a relatively comprehensive introduction to the device, giving the reader an intuitive feel for the relationship between the transducer overlap width W , coupling gap G , and the mode structure. However, they have not published an accurate means of predicting the frequencies of the resonant and spurious modes.

Gopani and Horne³ more recently have presented the frequency characteristics of various WGC resonators they have fabricated. Their results indicate they have been able to reduce the level of the spurious modes considerably but they provide no insights into their methods for achieving this result.

In the process of designing a WGC resonator for a specific application one is typically concerned about the center frequency and bandwidth of the device as well as the spurious modes of the device. These devices are quite complicated and a detailed understanding must employ both the coupling of modes (COM) analysis^{4,5} and the treatment of the device as a SAW structure supporting a collection of guided modes. Both of these analyses are key but in this paper we will focus principally on the SAW waveguide aspects of the WGC resonator. It is by the investigation of these modes that one can better understand the frequency response of the device and in particular the spectral location and strength of spurious modes.

For this investigation a number of WGC resonator

devices were designed and fabricated each with different values for W , G , and cavity length C . Numerous measurements were made to determine the relationship between device geometry and the device transfer function but herein we will only discuss the amplitude and phase profiles of the SAW transverse modes which were measured using a knife-edge laser probe.⁶ In general, our approach was to measure the transverse mode profiles at the resonant frequencies associated with peaks in the magnitude response of the transfer function of the WGC resonator.

In this paper we predict the transverse beam profiles by two distinctly different techniques. These methods yield the eigenmodes of the device structure and they compare favorably in shape with the mode profiles measured using the laser probe. The comparison is not precise, however, primarily because our predictions represent the eigenmodes of the structure and our laser probe measurements indicate the SAW profile at a given frequency for a source condition not considered in either of the theoretical treatments. Even so the corroboration between theory and experiment represents a considerable increase in the collective understanding of WGC resonators.

In Sec. II we discuss the physical layout of the device and present a representative frequency response for the device. In Sec. III we present both of our methods for calculating eigenmodes for this structure. And finally in Sec. IV we present our experimental results and compare these with our theoretical predictions.

II. DEVICE DESCRIPTION

Our design for the WGC resonator is shown schematically in Fig. 1. The device can be thought of as two one-port SAW resonators coupled acoustically by the overlap of their respective eigenmodes. To first order the device operates in the following manner. The input signal excites the surface acoustic wave in the upper resonator and the acoustic energy that couples into the lower resonator is then converted to an output electrical signal by the piezoelectric action of the substrate material. A more detailed and thorough description of the device operation requires a

^{a)}Bell Northern Research Ltd., P.O. Box 3511, Station C, Ottawa, Ontario K1Y 4H7, Canada.

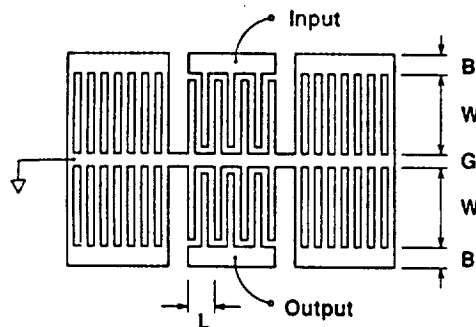


FIG. 1. Schematic layout for a waveguide-coupled resonator.

COM analysis^{4,5} which incorporates the currents and voltages of the input and output transducers as well as the forward and backward surface acoustic waves in the upper and lower resonators. One of the important parameters in the COM analysis is the coupling factor which describes how the SAW energy in one resonator couples to the SAW energy in the other resonator. The acoustic portion of this coupling would, in the present case, be related to the overlap integral of the eigenmodes of the two resonators. We will not in this paper be treating the COM analysis but will instead be laying the foundation for such a treatment by investigating thoroughly the eigenmodes of the WGC resonator.

In Fig. 2 we present the transfer function for a typical WGC resonator and it will be shown by theory and experiment in this paper that the eigenmodes play a vital role in determining the characteristics of the device transfer function. The passband response as well as the spurious response are set by the mode structure within the device. Peaks in the frequency response occur near frequencies which support both longitudinal and transverse modes. As will be shown, the lower edge of the passband f_{s1} in Fig. 2 is a symmetric mode and the upper edge of the passband

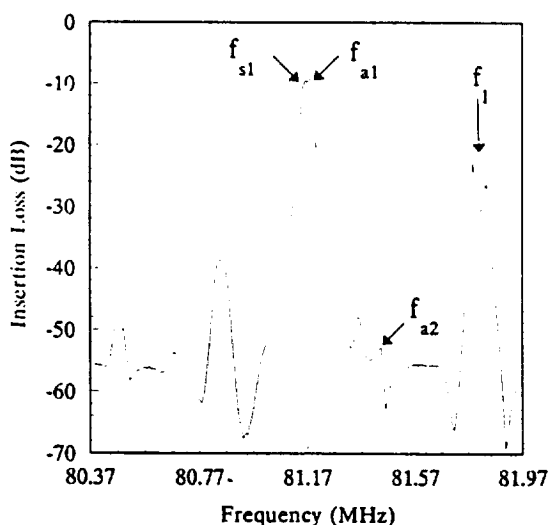


FIG. 2. Transfer function for a WGC resonator.

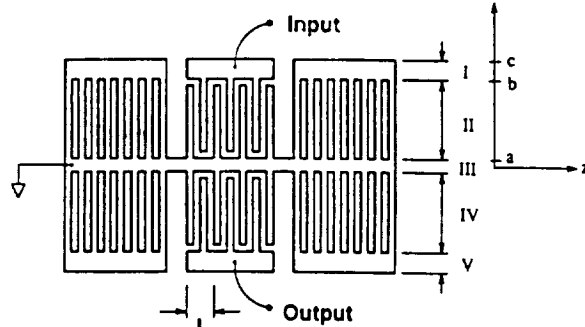


FIG. 3. WGC resonator as an array of five contiguous SAW propagation strips.

f_{a1} is an antisymmetric mode. The nearby spurious mode f_{a2} is a higher-order antisymmetric mode and at about the more distant frequency f_1 , the mode profiles associated with f_{s1} , f_{a1} , and f_{a2} are replicated about the next longitudinal mode frequency. The synthesis task which presents itself to the designer is to control the strength and location of these modes through design. The present work is a step towards this goal in as much as it provides some of the tools necessary to effectively perform this synthesis. Device dimensions and metallization thickness represent the principal design choices which must be made.

III. METHODS FOR EIGENMODE CALCULATIONS

In the following two subsections we present two independent methods for predicting the mode profiles in the WGC resonator. Both are later compared with the experimental results.

A. Analytical prediction of eigenmode structure

In order to analyze guided modes in a SAW waveguide, Knowles' scalar potential has been most widely used due to its simplicity.⁷ According to Knowles' analysis, the normal component of the displacement at the surface is proportional to a scalar potential ϕ for a semi-infinite isotropic solid.⁸ Moreover, this component of displacement can be directly measured by the knife-edge laser probe which was used for the experimental work presented later in this paper. Knowles' scalar potential will be used for the analytical approach as well as the stack matrix approach presented in the next subsection. The analytical approach is essentially a boundary value problem wherein the Helmholtz equation is solved for the scalar potential in the regions of different velocity within the waveguide structure and the boundary conditions of continuity are satisfied at each lateral interface. This approach is not as versatile as the stack matrix approach presented later.

We begin by thinking of the WGC resonator as nothing more than an array of five contiguous strips—each having a different velocity as shown in Fig. 3. Observe that we have, in the present analysis, only two distinct velocities. It should also be noted that in this paper we will not be addressing the issue of reflections within the device

structure, nor will we be considering the source conditions associated with the transducer SAW generation. The latter, in particular, will be the subject of future research.

Once we have segmented the WGC resonator into the five strips along the z direction of uniform velocity we proceed by solving the Helmholtz equation for the scalar potential ϕ ,

$$\begin{aligned}\frac{\partial^2 \phi_s}{\partial x^2} - \beta^2 \phi_s + \left(\frac{\omega}{V_s}\right)^2 \phi_s &= 0, \\ \frac{\partial^2 \phi_m}{\partial x^2} - \beta^2 \phi_m + \left(\frac{\omega}{V_m}\right)^2 \phi_m &= 0,\end{aligned}\quad (1)$$

where ϕ_s = the scalar potential for the semimetallized region, ϕ_m = the scalar potential for the metallized region, V_s = the SAW velocity in the semimetallized region, V_m = the SAW velocity in the metallized region, and ω = angular frequency.

We have assumed the z dependence for these scalar potentials to be of the form $e^{\pm j(\beta z - \omega t)}$, i.e., we are assuming the SAW propagation to be predominantly along the long axis of the WGC resonator. Under this assumption, the scalar potentials in each of the velocity regions of the resonator will be

$$\begin{aligned}\phi_I &= D_1 \sinh(k'_m x) + D_2 \cosh(k'_m x), \\ \phi_{II} &= D_3 \sin(k'_s x) + D_4 \cos(k'_s x), \\ \phi_{III} &= D_5 \sinh(k'_m x) + D_6 \cosh(k'_m x), \\ \phi_{IV} &= D_7 \sin(k'_s x) + D_8 \cos(k'_s x), \\ \phi_V &= D_9 \sinh(k'_m x) + D_{10} \cosh(k'_m x),\end{aligned}\quad (2)$$

where the term $e^{j(\omega t \pm \beta z)}$ has been dropped for clarity. Further, we have from the Helmholtz equation above that

$$\begin{aligned}k_m'^2 &= \beta^2 - k_s'^2, \quad k_s'^2 = k_s'^2 - \beta^2, \\ k_m &= \frac{\omega}{V_m}, \quad k_s = \frac{\omega}{V_s}, \quad \beta = \frac{\omega}{V_{\text{mode}}}.\end{aligned}\quad (3)$$

By applying the boundary conditions between the various regions of the structure that the scalar potential and its first derivative with respect to x are continuous, we can arrive at a dispersion equation which will enable us to compute the mode velocity V_{mode} . In addition we must recognize that there are symmetric and antisymmetric solutions to Eq. (1) and we must make judicious choices about the coefficients in Eq. (2) which can be set to zero.

Let us consider first the symmetric case. The requirement of symmetry about $x=0$ will force D_5 to be zero and hence we will have the symmetric form for the scalar potential in region III, i.e., $\phi_{III} = D_6 \cosh(k'_m x)$. For consistency and simplicity we take the form of the scalar potential in region II to be $\phi_{II} = D'_4 \cos(k'_s x + \theta)$. Applying the continuity boundary conditions at $x=a$ we will have

$$\begin{aligned}D_6 \cosh(k'_m a) &= D'_4 \cos(k'_s a + \theta), \\ D_6 k'_m \sinh(k'_m a) &= -D'_4 k'_s \sin(k'_s a + \theta),\end{aligned}\quad (4)$$

and by taking the ratio of these two equations we can eliminate the coefficients and have

$$\begin{aligned}k'_m \tanh(k'_m a) &= -k'_s \tan(k'_s a + \theta), \\ \theta &= \tan^{-1} \left(-\frac{k'_m}{k'_s} \tanh(k'_m a) \right) - k'_s a.\end{aligned}\quad (5)$$

In region I we can assume that the bus bar is wide enough that the scalar potential will go to zero at $x=c$. As such we can represent the scalar potential in that region as $\phi_I = D'_2 \sinh(k'_m(x-c))$. Applying the boundary conditions at $x=b$ we then have

$$\begin{aligned}D'_4 \cos(k'_s b + \theta) &= D'_2 \sinh(k'_m(b-c)), \\ -D'_4 k'_s \sin(k'_s b + \theta) &= D'_2 k'_m \cosh(k'_m(b-c)),\end{aligned}\quad (6)$$

and by taking the ratio of these two equations and using the expression for θ presented in Eq. (5) we have

$$\begin{aligned}-k'_s \tan \left[k'_s(b-a) + \tan^{-1} \left(-\frac{k'_m}{k'_s} \tanh(k'_m a) \right) \right] \\ = k'_m \coth(k'_m(b-c)).\end{aligned}\quad (7)$$

This can be simplified to be of the form

$$\begin{aligned}k'_s(b-a) &= \tan^{-1} \left(\frac{k'_m}{k'_s} \coth(k'_m(c-b)) \right) \\ &+ \tan^{-1} \left(\frac{k'_m}{k'_s} \tanh(k'_m a) \right).\end{aligned}\quad (8)$$

This transcendental dispersion relation can be solved for the mode velocity V_{mode} , which can in turn be used to find the resonant frequency for this mode given by $f = V_{\text{mode}}/2L$, where L is the grating period within the resonator defined in Fig. 1. The resonant frequencies associated with the symmetric modes will be represented as f_{sn} . Once we have the mode velocity we can sketch the mode profiles from the scalar potentials.

For the antisymmetric case we will take the odd form of the scalar potential in region III, i.e., $\phi_{III} = D_5 \sinh(k'_m x)$ and for simplicity take the form of the scalar potential in region II to be $\phi_{II} = D'_4 \cos(k'_s x + \theta)$. Again in region I we will take the scalar potential to be $\phi_I = D_1 \sinh[k'_m(x-c)]$. By proceeding in a manner similar to that presented above for the symmetric case we can obtain the dispersion relation

$$\begin{aligned}k'_s(b-a) &= \tan^{-1} \left(\frac{k'_m}{k'_s} \coth[k'_m(b-c)] \right) \\ &+ \tan^{-1} \left(\frac{k'_m}{k'_s} \coth(k'_m a) \right).\end{aligned}\quad (9)$$

In the same manner as was described for the symmetric case we can obtain the resonant frequencies f_{an} associated with the antisymmetric modes as well as the attendant mode profile. A comparison between the theoretical predictions and the experimental results will be presented in Sec. IV.

FIG. 4. N
rameters (

And
the proc
between th
and the
for a, b ,
in the v
dispersio
solved t
both the
frequen
antisymm
however
and exp
incorp
leads to
differen
velocity
the vel
V(
where
is the
pure m
algebra
portion

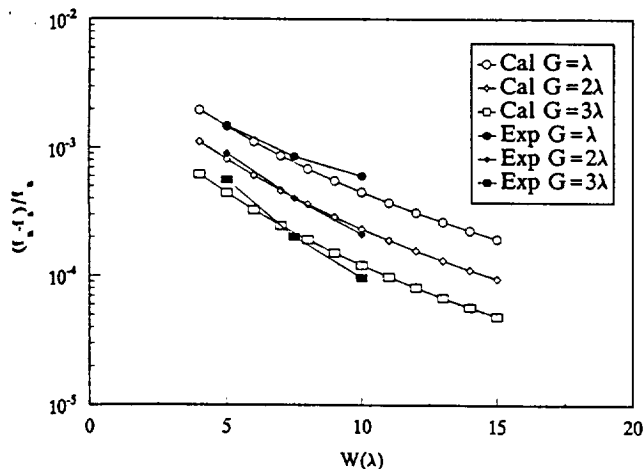


FIG. 4. Normalized frequency difference as a function of resonator parameters (W and G).

Another product of the analysis which is important in the process of filter synthesis involves the relationship between the symmetric and antisymmetric mode frequencies and the WGC resonator design features such as the values for a , b , and c shown in Fig. 3 as well as the SAW velocity in the various strips. For a given set of design criteria the dispersion relations represented by Eqs. (8) and (9) can be solved to find these frequencies. Tanaka *et al.*² presented both theoretical and experimental data for the normalized frequency difference between the first symmetric and first antisymmetric modes for a group of WGC resonators; however, there was not good agreement between theory and experiment. Upon further investigation we found that incorporation of the anisotropic nature of the substrate leads to faithful predictions of this normalized frequency difference. If the conventional assumption is made that the velocity anisotropy can be approximated by a parabola,¹¹ the velocity profile is represented by

$$V(\theta) = V_0 \left(1 + \alpha \frac{\theta^2}{2} \right), \quad (10)$$

where V_0 is the velocity along a pure mode axis and $V(\theta)$ is the velocity along some direction at an angle θ from the pure mode axis, and α is the anisotropy factor. After some algebraic manipulation¹² we can represent the appropriate portions of Eq. (3) as Eq. (3) can be rewritten as

$$\begin{aligned} k_m'^2 &= \beta^2 - (1 + \alpha_m) k_s^2, \\ k_s'^2 &= (1 + \alpha_s) k_s^2 - \beta^2, \end{aligned} \quad (11)$$

where α_m and α_s represent the anisotropy coefficient for the metallized and semimetallized regions, respectively. For simplicity we have used the value of the anisotropy coefficient for STX quartz, $\alpha = 0.378$, for both α_m and α_s . The relations presented in Eq. (11) were then used in Eqs. (8) and (9) and the normalized frequency difference was computed for several values of W and G . The results of these calculations are presented in Fig. 4 along with the experimental data of Tanaka *et al.*² The inclusion of the anisotropy

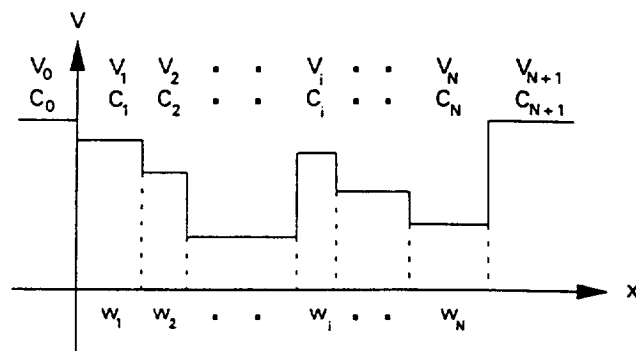


FIG. 5. SAW velocity profile in an arbitrary N -multichannel waveguide, where V_i is the velocity and C_i is the stiffness constant in the i th region, respectively. The width of the i th channel is given as W_i .

factor brings the theory in very close agreement with the experimental results—much closer than when the anisotropy is not included.

B. Stack matrix theory

The transverse profile of the SAW velocity in a general N multichannel waveguide structure is schematically shown in Fig. 5. The coordinate system is chosen so that the x axis is along the transverse direction and the z axis is along the propagation direction. The wave is guided in the slow velocity sections by the nature of the waveguide. In order to analyze guided modes in a SAW waveguide, we will use the Knowles' scalar potential used for the work presented in Sec. II A.

The propagation equation of the scalar potential ϕ_i , at the surface in section i , is given by a second-order ordinary differential equation as follows:

$$\frac{d^2 \phi_i(x)}{dx^2} + \left(\frac{\omega^2}{V_i^2} - \beta^2 \right) \phi_i(x) = 0, \quad (12)$$

where β is the longitudinal propagation constant, ω is the angular frequency, and $V_i(x)$ is the SAW velocity in the i th section.

A general solution of this equation is

$$\phi_i(x) = A e^{-jk_i x} + B e^{jk_i x}, \quad (13)$$

where $k_i = \sqrt{\omega^2/V_i^2 - \beta^2}$. We have made the conventional assumption that the z dependence is $e^{-j\beta z}$ and the time dependence is $e^{j\omega t}$.

The boundary conditions at the interface between the i th and the $(i+1)$ th section shown in Fig. 5 are that the displacement and the stress are continuous. These boundary conditions can be denoted by

$$\begin{aligned} \phi_i|_{x_i} &= \phi_{i+1}|_{x_i}, \\ C_i \phi_i'|_{x_i} &= C_{i+1} \phi_{i+1}'|_{x_i}, \end{aligned} \quad (14)$$

where C_i is an effective stiffness constant of the i th section and the prime denotes differentiation with respect to x .

The close analogy between the scalar potential in SAW and the electric field in optical wave makes it possible to

use theories previously developed for the optical waveguides. Among various techniques for calculating the fields in optical waveguides, stack matrix theory has been adapted herein due to its capacity for treating arbitrary structures. Only the key portion of the theory is described in this paper and those seeking more details are referred to Kogelnik and Shank's formulation.⁹

Stack matrix theory begins by defining the continuous components used in the boundary condition to be the field variables U and W as follows:

$$U_i = \phi_i(x) = A e^{-jk_i x} + B e^{jk_i x}, \quad (15)$$

$$W_i = C \phi'_i(x) = j k_i C_i (-A e^{-jk_i x} + B e^{jk_i x}).$$

After some algebraic manipulations, one can obtain a simple relationship between two adjacent field variables as follows:

$$\begin{bmatrix} U_{i-1} \\ W_{i-1} \end{bmatrix} = \begin{bmatrix} \cos k_i w_i & -\frac{1}{C_i k_i} \sin k_i w_i \\ C_i k_i \sin k_i w_i & \cos k_i w_i \end{bmatrix} \begin{bmatrix} U_i \\ W_i \end{bmatrix} = M_i \begin{bmatrix} U_i \\ W_i \end{bmatrix}, \quad (16)$$

where w_i is the width of the i th section and M_i is the *characteristic matrix* of the section. It should be noted that the x axis is shifted so that each interface position of the corresponding field variables is to become zero. The repeated application of the boundary conditions at each interface is accomplished by the concatenation of each M_i . Therefore, the *characteristic matrix* of the stack M , which includes the entire structure, is given by

$$\begin{aligned} \begin{bmatrix} U_0 \\ W_0 \end{bmatrix} &= M_1 M_2 \cdots M_N \begin{bmatrix} U_N \\ W_N \end{bmatrix} \\ &= M \begin{bmatrix} U_N \\ W_N \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} U_N \\ W_N \end{bmatrix}. \end{aligned} \quad (17)$$

For the guided wave solutions, ϕ should be evanescent at both ends, i.e., for both $i=0$ and $i=N$ sections. This guided condition is satisfied by a pure imaginary value of k_i at both sides and $B_0 = A_{N+1} = 0$. This can be expressed by the field variables

$$\begin{bmatrix} U_0 \\ W_0 \end{bmatrix} = \begin{bmatrix} A_0 \\ C_0 \gamma_0 A_0 \end{bmatrix} \quad (18)$$

and

$$\begin{bmatrix} U_N \\ W_N \end{bmatrix} = \begin{bmatrix} B_{N+1} \\ -C_{N+1} \gamma_{N+1} B_{N+1} \end{bmatrix}, \quad (19)$$

where $\gamma_0 = j \sqrt{\beta^2 - (\omega^2/V_0^2)}$, $\gamma_{N+1} = j \sqrt{\beta^2 - (\omega^2/V_{N+1}^2)}$. Inserting these expressions into the stack matrix relation, one can obtain

$$\begin{bmatrix} A_0 \\ C_0 \gamma_0 A_0 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} B_{N+1} \\ -C_{N+1} \gamma_{N+1} B_{N+1} \end{bmatrix}. \quad (20)$$

A rearrangement leads to a matrix equation

$$\begin{bmatrix} 1 & -m_{11} + C_{N+1} \gamma_{N+1} m_{12} \\ C_0 \gamma_0 & -m_{21} + C_{N+1} \gamma_{N+1} m_{22} \end{bmatrix} \begin{bmatrix} A_0 \\ B_{N+1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (21)$$

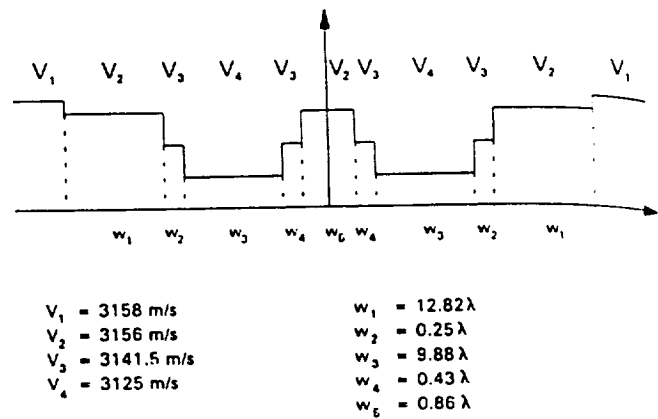


FIG. 6. SAW velocity profile for our WGC resonator.

In order to obtain a nontrivial solution, the determinant of the matrix must be set to zero thus yielding the dispersion relation for the waveguide. For the general multichannel waveguide structures, a numerical method is necessary to solve the matrix equation.

The velocity profile representative of the WGC resonator is shown in Fig. 6. The velocities presented in this figure are V_1 =free surface velocity on STX quartz, V_2 =velocity underneath the aluminum bus bars, V_3 =velocity in the region between the ends of the IDT electrodes and the bus bars, V_4 =velocity in the region of overlap between the IDT fingers and we have used standard methods¹³ to compute the values. If we assume the velocities are isotropic, we can obtain the maximum angle the acoustic energy deviates from the longitudinal direction to be $\theta = \cos^{-1}(V_4/V_1)$, where V_4 and V_1 are the slowest and the fastest velocities, respectively. Using the values of 3125 and 3158 m/s for V_4 and V_1 , respectively, we can expect the angle to be about 8.3° . If we now consider the true anisotropic nature of the velocity on STX quartz we see that the velocity change over this 8.3° is only $\sim 0.3\%$.¹⁰ Therefore, we can assume that the velocity is isotropic without appreciably affecting the predicted mode profiles. The effective stiffness constants are used in the boundary condition presented in Eq. (3) are assumed to be the same because each section is elastically very similar in this kind of low dispersion waveguide.

IV. MEASUREMENTS AND RESULTS

In this section we will present experimental measurements of a WGC resonator on STX quartz which had an electrode thickness to SAW wavelength ratio of $\sim 2\%$. We will compare these results with theoretical predictions based on the methods presented in Sec. III. Though numerous devices were fabricated and tested we will present only the results from one of these which had strong spurious modes. While this device would not be good for a system application, it was an ideal device for our mode profile measurements. The mode profiles we present for this one device are indeed representative of the profiles we measured for a large number of devices.

The device was mounted on the computer-controlled translation stage of our knife-edge laser probe and leveled. The details of operation of this apparatus have been described previously⁶ and will not be repeated here. It is sufficient to say that the laser probe allows us to measure the magnitude and phase of a Rayleigh wave at various locations throughout the device. For these experiments the laser spot at the surface of the device was $\sim 4 \mu\text{m}$ in diameter and the signal-to-noise ratio of the measurements was typically 60 dB. The successful operation of the laser probe requires that the location of the scan on the surface of the device have a reasonable reflectance. It is difficult to obtain a strong laser probe signal on a translucent material such as quartz and so the measurements were made with the laser spot on the aluminum electrodes of the interdigitated transducer (IDT) and the reflector grating rather than on the bare quartz gaps between the electrodes.

From the device transfer function shown in Fig. 2 we select the frequencies at which we wish to make transverse laser probe scans in the x direction as indicated in Fig. 1. For example, if we wish to measure the mode profile for the first symmetric mode we set the SAW drive for the device at the first symmetric mode frequency f_{s1} , and scan transverse to the direction of SAW propagation. Similarly we can select the frequencies associated with any of the spurious modes which appear in the transfer function and measure the associated mode profiles.

For the mode profile data we present the origin of the coordinate system was selected to be on the central bus bar in the center of the resonator cavity. Another way to look at this is to say that the origin was selected to be at the geometric center of the device. For the three mode profiles we show the scans were taken transverse to the predominant direction of SAW propagation at the center of the cavity. The positive values of the transverse dimension (x) as shown in Fig. 7 through 9 represent the track driven by the input signal and the negative values represent the output track. In order to obtain the symmetric mode profiles the device was driven symmetrically, i.e., the same signal was applied to the input and output transducers. For the antisymmetric mode profiles a balun was used so that the input and output transducers could be driven 180° out of phase.

In Fig. 7(a) we show the measured and predicted normalized magnitudes of the first symmetric mode which occurred at 81.153 MHz. The magnitude is expressed in arbitrary units (A.U.). Both theoretical approaches very closely predict the measured results with the profile reaching peaks in the middle of the input and output tracks and a local minimum on the grounded bus bar between. In Fig. 7(b) we present the phase for the first symmetric mode and see again that there is relatively close agreement between theory and experiment. It should be pointed out that the phase is indicative of the phasor relationship between the rf signal driving the SAW transducer as it leaves the network analyzer and the amplified photodetector signal received by the network analyzer. As such the phase, as is the case for the magnitude, can at best be regarded as a relative, not an absolute, measurement.

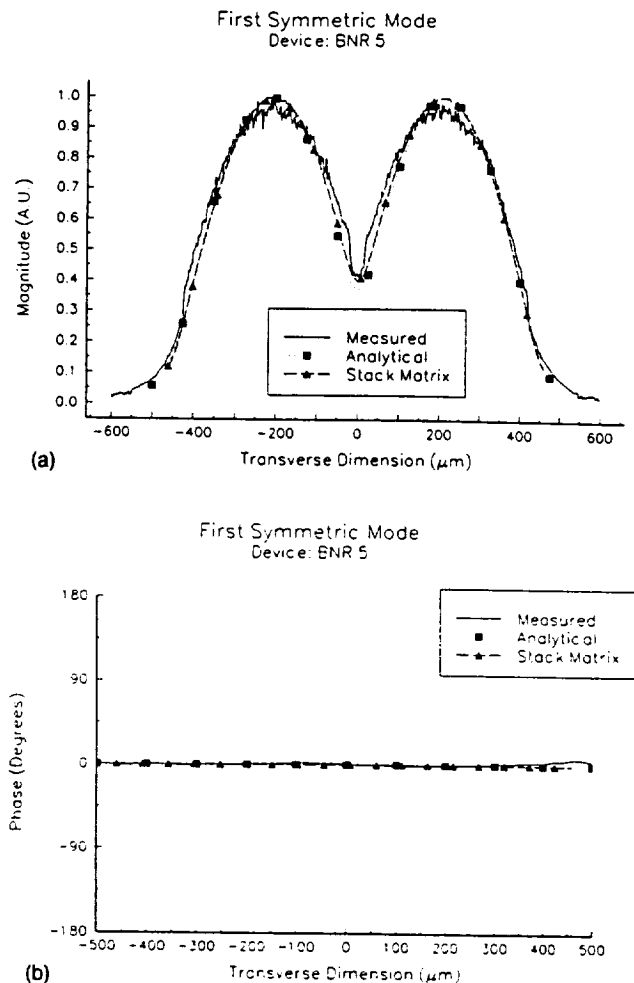


FIG. 7. Transverse beam profile for first symmetric mode (a) magnitude, (b) phase.

In Fig. 8 we present the magnitude and phase for the mode profile of the first antisymmetric mode which occurred at 81.1887 MHz. Again there is good comparison between theory and experiment for both the magnitude and phase. The magnitude profile, shown in Fig. 8(a), is of course quite similar to that for the first symmetric mode with the exception that the profile goes to zero at the central bus bar for the antisymmetric mode.

In Fig. 9 we present the magnitude and phase for the mode profile of the second antisymmetric mode which occurred at 81.446 MHz. As has been the case for the previous modes discussed there is good agreement between theory and experiment with regard to the overall shape of both the magnitude and the phase but not quite as good. Note in particular that the predicted nulls at the center of the transducer regions ($\sim \pm 220 \mu\text{m}$) do not occur for the measured profile and this can best be explained as follows. For our experiments with the laser probe we measure the SAW profile generated in the device at a single drive frequency under specific drive conditions. What we have found from experience is that we are measuring the sum of all of the mode profiles that are generated at this single frequency. For the second antisymmetric mode shown in Fig. 9 we drove the input and output transducers 180° out

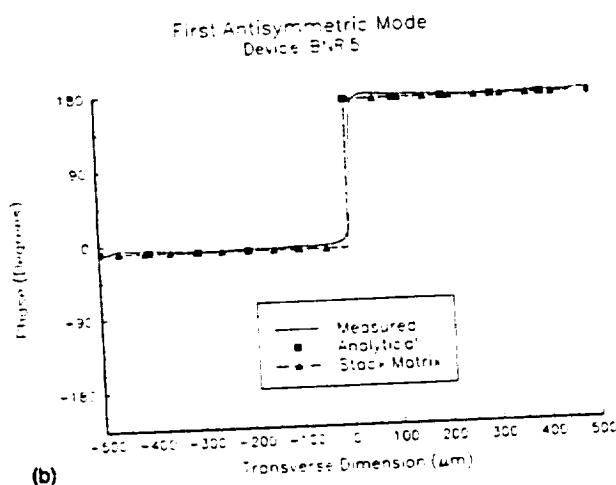
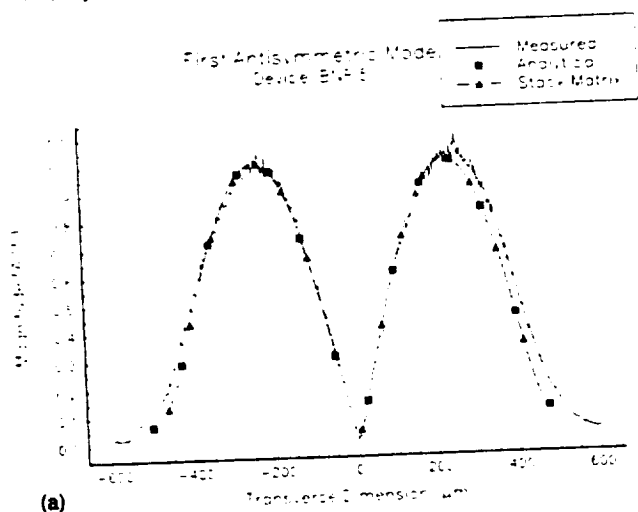


FIG. 8. Transverse beam profile for first antisymmetric mode (a) magnitude, (b) phase.

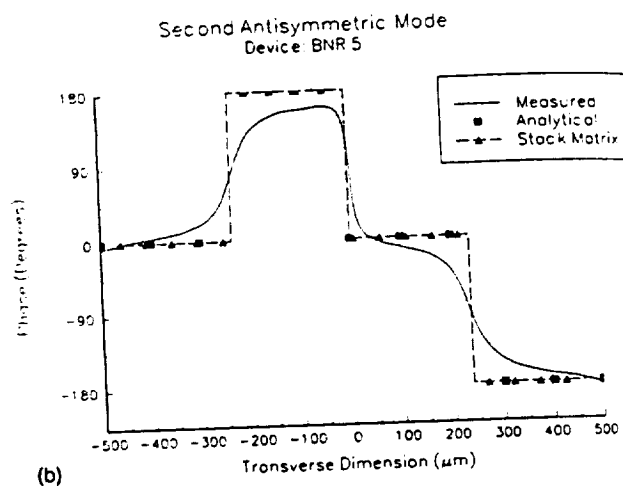
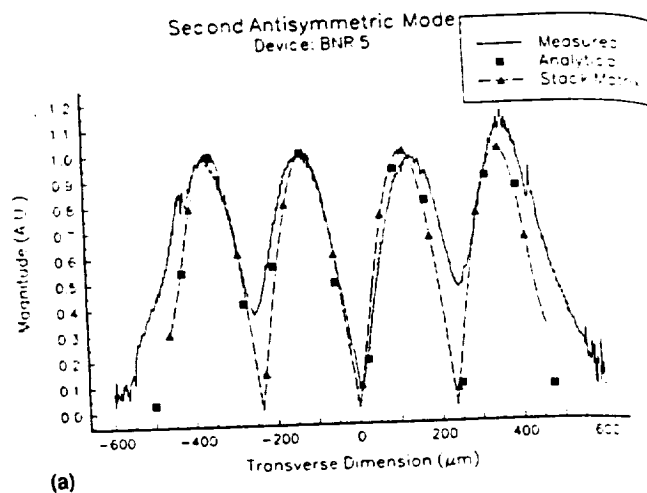


FIG. 9. Transverse beam profile for second antisymmetric mode (a) magnitude, (b) phase.

of phase and as such generated a number of antisymmetric modes. Predominant among these are the second antisymmetric mode and a weakened first antisymmetric mode. One can then view the absence of the deep nulls predicted for Fig. 9 as the sum of the nearly zero amplitude for the second antisymmetric mode with the peak amplitude of the first antisymmetric mode.

V. SUMMARY AND CONCLUSIONS

In this paper we have presented both experimental and theoretical work on the mode profiles in WGC resonators. The two independent methods for predicting the mode profiles are reasonably accurate though the stack matrix theory method is the most versatile of the two. This paper represents, to the best of our knowledge, the first use of stack matrix theory to predict mode profiles in SAW structures. The versatility and accuracy of this technique makes it useful for the calculation of mode profiles in complicated, multichannel SAW and ACT devices. In addition, the anisotropic nature of the SAW velocity was included yielding a more accurate prediction of the normalized frequency difference and hence a superior filter design tool.

ACKNOWLEDGMENTS

This work has been supported by Bell Northern Research Ltd. (BNR), the National Science Foundation through a Presidential Young Investigator award, and by NASA under Grant No. NAGW-2753. The assistance of these organizations is gratefully acknowledged. We also wish to thank Grantley Este of BNR for many helpful discussions with regard to the fabrication of SAW devices and Conrad Gratton of BNR for initial testing and packaging of the devices.

- ¹A. Yamada and H. Shimizu, Paper of the Technical Group on Ultrasonics, IECE, Japan, VS77-33, 1977, p. 29.
- ²M. Tanaka, T. Morita, K. Ono, and Y. Nakazawa, 38th Annual Frequency Control Symposium, 1984, pp. 286-293.
- ³S. Gopani and B. A. Horine, Proceedings of the 1990 IEEE Ultrasonics Symposium, 1990, pp. 1-5.
- ⁴H. Kogelnik and C. V. Shank, J. Appl. Phys. 43, 2327 (1972).
- ⁵H. A. Haus and P. V. Wright, Proceedings of the 1980 IEEE Ultrasonics Symposium, IEEE Cat. No. 80CH1602-2, 1980, Vol. 1, pp. 277-281.
- ⁶R. L. Miller, Ph.D. dissertation, University of Illinois, Urbana, Illinois, 1987.
- ⁷R. V. Schmidt and L. A. Coldren, IEEE Trans. Sonics Ultrasonics SU-22, 115 (1975).
- ⁸J. K. Knowles, J. Geophys. Res. 71, 1580 (1966).

- ⁹H. Kogelnik, in *Guided Wave Optoelectronics*, edited by T. Tamir (Springer, New York, 1988).
- ¹⁰B. K. Sinha and H. F. Tiersten, *J. Appl. Phys.* **52**, 7196 (1981).
- ¹¹T. L. Szabo and A. J. Slobodnik, *IEEE Trans. Sonics Ultrasonics* **SU-20**, 240 (1971).

- ¹²T. P. Cameron, M. S. thesis, Carleton University, Ottawa, Ontario, Canada, 1988.
- ¹³S. Urabe, Y. Koyamada, and S. Yoshikawa, *Trans. IECE Jpn.* **J60-A**, 875 (1977).

5.5 Suggestions for Future Work:

We have had some recent inquiries regarding ACT device work and it is felt that a first step would be a study to see if one of the existing GaAs device architectures for commercially available RFICs could be used for ACT devices. If the ACT device can be made to piggyback on an existing process then we believe the prospects are very good that cost-effective devices can be manufactured. It would also be advisable to use the IDT-less ACT approach which we have developed along with a ZnO overlay. In this manner the imager chip could be driven by a modest voltage and power with no need for timing circuitry. Also, if the imager requires too large a capital investment in light of the (now) current availability of Silicon CCD imagers for HDTV, the ACT technology may provide a good way to realize a high speed image processing chip for HDTV. Prior to digitization, the ACT could be used for on chip image enhancement which could process data in both space and time.

SECTION VI:

KODAK

HDTV Prototype Imaging System

Introduction

The HDTV imaging system is a prototype system and therefore not intended for scientific imaging applications. Like all prototypes, it has many limitations which affect image quality and hardware reliability. This prototype camera features the Kodak KAI-2091M high-definition monochrome image sensor. This progressive scanned charge-coupled device (CCD) can operate at video frame rates and has 9 micron square pixels for easy image processing. The photosensitive area has a 16:9 aspect ratio and is consistent with the "Common Image Format" (CIF). It features an active image area of 1928 horizontal X 1084 vertical pixels and has a 55% fill factor. The camera is designed to operate in continuous mode with an output data rate of 5MHz, which gives a maximum frame rate of 4 fps. Due to frame buffer limitations of the ITI imaging system, only the first 1024 lines can be processed and displayed.

Imaging System Setup and Operation

To setup the ITI imaging system, connect the two Sony monitors, keyboard, and mouse to the ITI computer. There are two monitor connectors located on the back of the computer. Either monitor can be attached to either monitor connector. One monitor is used to view images while the other is used to operate the computer. A power cord for each monitor and the computer also need to be connected. To setup the camera, attach the Nikor zoom lens (70-210mm) to the front of the camera. Plug the camera power supply into a standard wall outlet. Make sure the camera supply is turned off. Connect the camera power supply cable to the camera using the 9 pin pigtail connector. Connect the 68 pin SCSI connector from the camera to the ITI imaging system. Apply power to the camera by turning on the camera power supply. Apply power to the ITI imaging system. Once the

ITI system has booted, choose the HDTV icon to operate the camera. To obtain the best possible picture, adjust the lens to control the amount of light entering the camera. This can be accomplished by opening and closing the lens aperture.

Power Requirements

A power box containing several linear commercial power supplies provides the power to run the camera head through a 9 pin pigtail connector. The power box operates off a standard 115VAC 3 prong wall socket and is fused to protect the internal supplies according to manufacturer specifications. The current draw from each internal linear supply is shown below. The .5 amps @ +5VDC required to power the termination resistors for the 68 pin SCSI cable is included in the totals below.

1.35 amps @ +5 volts DC

.12 amps @ +15 volts DC

.08 amps @ -15 volts DC

.05 amps @ +24 volts DC

To ensure the correct polarity for all DC operating voltages, use the power box and cable provided with the camera. If a different power supply is used, the camera may be permanently damaged if the DC voltage polarities are reversed.

Prototype Camera Performance

Black Level	Adjusted to no light condition
Scanning	Non-Interlaced, progressive
Synchronization	Internal pixel clock, line sync, frame sync
Resolution	8 bits
Pixel Clock Rate	5MHz
Frame Rate	4 fps

Interface Specifications

The AIA interface connector is a 68 pin, high density, dual row, D-type connector. The connector has .050 pin spacing and a D-type shell that is 2.5 inches long. The connector carries the signals needed to interface the camera to the ITI frame grabber. The signals include the two digital video output channels, A and B, each with eight bits labeled DATA0 through DATA7. DATA0 is the least significant bit and DATA7 is the most significant bit. In addition, there are three timing output signals which control the transfer of video data. They are the FRAME ENABLE, LINE ENABLE, and PIXEL DATA STROBE. All signals are output as differential pairs with signal levels conforming to the RS422 specification. The noninverting part of the differential pair is present on the (+) output, while the inverting part of the differential pair is present on the (-) output. The chart below lists each signal name and its pin number.

68 pin camera connectorSignal Name44 pin ITI connector

9	ADATA 0+	1
43	ADATA 0-	16
8	ADATA 1+	31
42	ADATA 1-	2
7	ADATA 2+	17
41	ADATA 2-	32
6	ADATA 3+	3
40	ADATA 3-	18
5	ADATA 4+	33
39	ADATA 4-	4
4	ADATA 5+	19
38	ADATA 5-	34
3	ADATA 6+	5
37	ADATA 6-	20
2	ADATA 7+	35
36	ADATA 7-	6
20	BDATA 0+	21
54	BDATA 0-	36
19	BDATA 1+	7
53	BDATA 1-	22
16	BDATA 2+	37
50	BDATA 2-	8
15	BDATA 3+	23
49	BDATA 3-	38
14	BDATA 4+	9
48	BDATA 4-	24
13	BDATA 5+	39
47	BDATA 5-	10
11	BDATA 6+	25
45	BDATA 6-	40
10	BDATA 7+	11
44	BDATA 7-	26
25	FRAME ENABLE+	12
59	FRAME ENABLE-	27
26	LINE ENABLE+	14
60	LINE ENABLE-	29
29	PIXEL DATA STROBE+	28
63	PIXEL DATA STROBE-	42
30	EXP+	41
64	EXP-	13
1 & 35	GND	44

Prototype Camera Head Design

The HDTV camera head design is partially based on the MITE camera head design. The electronics has been modified to meet the requirements of the HDTV sensor and is designed for maximum flexibility and testability. All clock drivers and biases can be adjusted without changing components. Sensor timing is generated using one EPLD. All EPLD outputs are routed to diagnostic connectors which interface directly to a logic analyzer. This allows timing performance to be monitored while viewing the image data.

All sensor timing is based on a 80MHz crystal which feeds the global clock input to each EPLD. The 80MHz clock provides a minimum time resolution of 12.5ns. Code was then developed using Altera's Max PLUS II software to create the clock signals required by the sensor. The sensor is currently being clocked at 1/16 the crystal frequency or 5MHz. This provides a camera frame rate of approximately 4 frames/sec.

All timing logic was simulated using the Altera Max Plus II software. Each simulation allows you to verify your logic by displaying the timing visually. If the simulator shows a logic problem, you can easily modify the timing equations, recompile the program, and simulate the waveforms again. A number of simulations were performed using various clocking scenarios to determine optimal performance for this system.

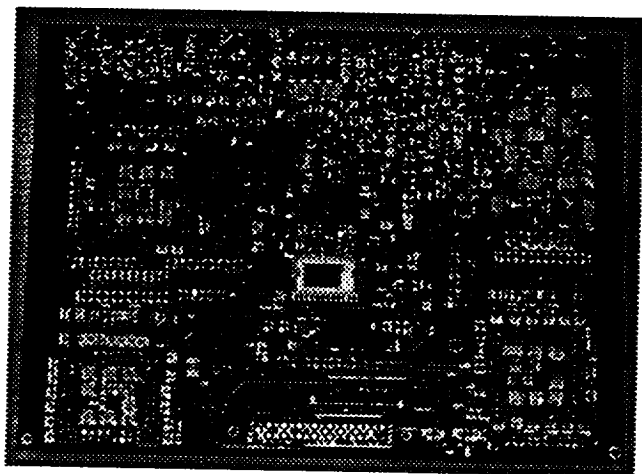
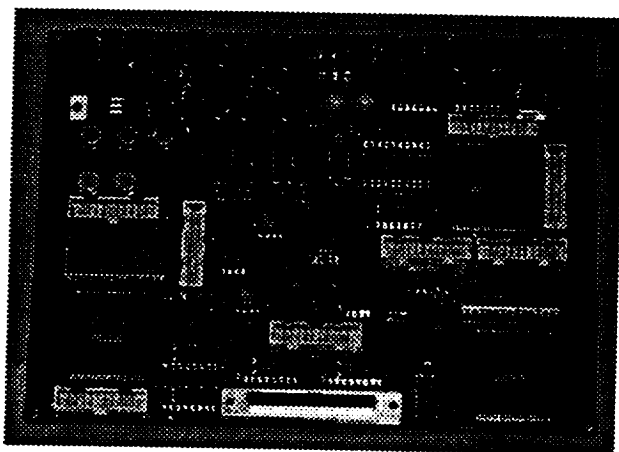
The camera head electronics was packaged to fit on one 6 x 8 circuit board to eliminate sending the 80MHz clock signal between boards. The three EPLD's are co-located to minimize the distance to the 80MHz crystal. Two EPLD's perform the digital clamp and sample for video channel A and B. The third EPLD generates timing signals for the sensor and frame grabber interface.

Since there are two output video channels, the gain and offset of each channel must be closely matched to minimize line to line shading. Both these parameters are factory set. The black level can be monitored by viewing the video output with the lens capped. The gain is set so the video output signal is at maximum amplitude when the detector reaches saturation.

System Level Integration

To view HDTV images, the Camera Head was integrated with an ITI frame grabber and image display system. The ITI system consists of a PC with ITI image processing boards and software. Application code was developed using the ITI software development tools to acquire, process, and display images from the camera head via an RS 422 interface. The IM-PCI image processing card can handle two 20MHz data channels which correspond with the two video taps from the camera head. This allows the ITI system to theoretically handle and display image data at 15 frames/sec.

The ITI imaging system is packaged in a standard 19" rack mountable industrial chassis. It uses a 133MHz Intel Pentium Processor with 32 MB of memory and runs the Windows 95 operating system. Two 17" Sony Trinitron monitors are used with the system. One monitor operates the computer and the other displays images. The computer has one full-size single PCI bus card to manage memory, one computational module for real-time statistical analysis, and one digital acquisition module to interface with the camera.



SECTION VII:

MIT/POLAROID

Memo to: Bill Hunt
From: Richard Solomon and Clark Johnson, University of Pennsylvania
Date: 19 September 1999
Re: MIT/Polaroid Camera Report

Bill:

This is a response to your e-mail message of 18 August. Under separate cover, by Priority Mail, we are sending a number of support documents. These include the camera specifications, a reprint on the actual camera system operational details, along with two internal MIT memos that might be of interest to you. In addition, we are sending a CD-ROM (see below) of some images produced by the Navy.

There is no operations manual specifically for this camera. Philips did supply a generic manual for high-end video cameras. However, the various agencies that now have cameras readily taught themselves effective operations with technical help from Philips and Polaroid personnel.

The first two cameras came with only one Fujinon zoom lens that cost in excess of \$100,000. It turned out to be inappropriate for certain scientific and cinemagraphic applications. For these uses a special mount was made for modified Hasselblad 2 1/4 square (SLR) and conventional motion picture camera lenses.

Both of these cameras are the property of NASA, under the supervision of Dr. Tice DeYoung, the original project initiator when he was at DARPA. One of the cameras is currently assigned to the supercomputer center at the Naval Research Laboratory in Anacostia, where it is available for demonstration.

The other camera is assigned to the National Security Agency and is on temporary loan to Demografix in Santa Monica, CA. Four additional cameras have subsequently been purchased and delivered to the Navy for further research. One of the original cameras may be transferred to the University of Pennsylvania shortly. The University of Pennsylvania is currently under contract with the Naval Research Lab to continue experimentation under a program on Vision Science and Advanced Networking. Principals include Prof. David J. Farber and the authors of this report.

Additional cameras have been purchased by ABC/Disney and other commercial firms. Demonstrations have been made at the National Institutes of Health, the U. S. Army, the National Security Agency, the National Association of Broadcasters annual meeting in Las Vegas (where it won Best of Show), and at the American Society of Cinematographers' annual convention in Los Angeles. The camera was featured in a weekly column on advanced multimedia technology in the New York Times shortly after it was delivered by Polaroid.

A camera was loaned to ABC/Disney for a series of test shoots at various sites across the country. These shoots provided material for A/B comparison testing with so-called HDTV interlaced cameras. The Navy did airborne tests from a dirigible of the Norfolk (VA) Navy Yard and a beach. A few freeze frames showing the spectacular results of

this Navy shoot have been captured on a CD-ROM, a copy of which is being sent to you. (Copies are available upon request.)

In the summer of 1998 a camera was loaned to Mr. Robert Primes (an award-winning cinematographer, and a former president of the ASC), and Gary Demos of Demografix (an Oscar-winning special-effects expert) who together financed the production video of Mrs. Primes (a professional concert pianist) performing a Chopin etude. This astonishing 7-minute movie can only be fully appreciated on a specially modified projector for an 18-ft. screen, one of which is at NRL. Editing was an exceptionally time consuming exercise and had to be done frame by frame on special Apple Macintosh computers.

As an example of this remarkable presentation, one can readily see the vibrating piano strings without a scintilla of blur or interlace artifacts. It is as if one is sitting next to the pianist. The Hasselblad lens was critical for the close-ups in this shoot.

The entire high-resolution system (camera and display) has been successfully tested in real time on the high-bandwidth government-operated ATDnet fiber-optic research network in the Washington/Baltimore area over a simulated distance of several hundred miles.

The only current storage device capable of handling the high bit rate generated by the camera without distortion-creating compression is a Matsushita D-5 video recorder modified to handle a linear bit stream. The Naval Research Laboratory recently ordered a RAID array capable of storing 20 minutes of uncompressed camera output. One camera has been returned to Philips to be modified to operate at 72 frames/second.

Future plans include building a test-bed camera capable of:

1. Variable frame rate (24-72 f/sec.);
2. Wide color gamut using additional color sensors and variable color filters;
3. Smaller CCD target to permit use of conventional high-quality 35mm still-camera lenses;
4. Transcontinental testing on the gigabit optical network with analog transmission and digital processing at the display;
5. Variable spatial and temporal modulation transfer functions;
6. Gigabit local rf transmission in a local environment.

All of the participants involved consider this to be a most successful demonstration of the performance advantages of progressive imaging video technology.

Demonstrations on the giant screens using the Polaroid/MIT camera and the special high-resolution projector can be arranged at either the Naval Research Laboratory in Anacostia, or in Santa Monica, CA.

Capturing Reality: A New Full-Motion Electronic Camera

MIT AND POLAROID CORP. have developed a high-resolution camera which demonstrates a full-motion electronic imaging system that comes closest to representing reality.

The camera does this by mimicking in various respects the way that the eye and brain capture and partially process real life images.

There are some 40 million high-resolution desktop computer monitors in the U.S. today which can immediately display the output of this camera.

The particular design of the capture system enhances virtually lossless compression facilitating both realtime transmission over such advanced links as digital subscriber loops (XDSL telephone technology) or super-compressed storage media (DVD disks, for example). 200:1 compression in non-realtime, and 50:1 in realtime, has been demonstrated using progressive capture imaging.

The camera can be thought of as an electronic "Hasselblad_" for full-motion imaging. *It combines the best elements of film and electronics to produce a rock-steady, grainless, high-resolution, full-spectrum image with imperceptible scan lines and more than adequate light sensitivity.* The image displayed on conventional high-resolution computer monitors appears to the human eye what it would see if present at the scene. The camera does this by processing images differently than, and avoiding most of the defects of, existing television and video systems.

The camera is part of a modular systems design facilitating the maximum use of *uncoupled and interoperable basic computer devices* for display, transmission, processing, and storage. The implications for advanced electronic motion photography are obvious. The system has direct and immediate applications in telemedicine, teleconferencing, instrumentation, intelligence and surveillance missions.

Our system is the only one of the 17 formats recently adopted by the Federal Communications Commission for broadcast HDTV that is fully compatible with computer processes. The ability to combine real images with software-generated graphics for computer-assisted training, maintenance and medical procedures is part of our system's basic design architecture.

New imaging applications require a new video system

While conventional television techniques have been adequate for entertainment electronic imaging needs up to now, important new uses of video have emerged requiring an image fidelity well beyond the needs of broadcast television production.

The instrumentation, medical, training and scientific research establishments will likely be a much larger customer for reality-based advanced imaging devices than the entertainment industries. Industrial, scientific and medical photography, education and training, all require imaging technology without the discernible artifacts found in current, standard professional-quality video cameras. (Indeed, the illusions created by the visual escape from reality in current television and film-based motion pictures are often required, but are in contradiction to what is needed in non-entertainment electronic imaging.)

Interoperability with digital computers and advanced digital telecommunications systems, scalability upwards and downwards depending on applications, extensibility to new technology, and open architectures permitting user modifications are key characteristics of such an approach. Scientific and industrial requirements for imaging technology were precisely what DARPA had in mind when it contracted with MIT to determine the feasibility of developing new camera technology that uses coherent, progressive image capture at a high resolution.

How this new imaging system works

The initial development project designed a camera with a *CCD that captures the entire image simultaneously*, prior to electrical read out (similar to the rods and cones in the human eye which are sensitized in parallel). This CCD does not scan the image like all other commercial video and television cameras.

As proof of principle, two cameras meeting MIT's advanced performance specifications were built under contract with Polaroid. Polaroid subcontracted with Philips N.V. to adapt off-the-shelf high-end camera bodies with the improved CCD imaging arrays. The CCDs were manufactured by IBM from Polaroid masks.

The cameras have effectively four times current video spatial resolution and at 60 frames per second, double the normal video frame rate. But spatial and temporal resolution are not enough to create an illusion represented reality: *virtually noise-free, coherent, and progressive image capture* make this camera design different from all existing commercial video devices.

The camera maintains comparable sensitivity and color balance associated with conventional high quality professional cameras. Moreover, the image raster is square, and the aspect ratio can be made variable. This *facilitates superimposition of computer-generated images with real life images*, something almost impossible on current video systems with rectangular "pixels" and inflexible image aspect ratios.

Defects of current television

As the science of opthamology has long recognized, *the eye simply does not work like a photographic camera*, however much it superficially resembles such a device. Human sensitivity to video artifacts found in current systems is inherent in the extremely complex mechanism of the eye/brain linkage. We have demonstrated that artifacts can be largely eliminated by using different mechanism for image capture.

Several inherent characteristics of conventional television systems, including most proposed HDTV systems, make it practically impossible to capture, record and display electronic images without degrading artifacts. Television, as most of us know it, never quite looks like a real scene. Movies come closer, but film degrades images in different ways because of its random, bouncing grain patterns and frame registration ("judder") problems. Yet, both of these media may be quite adequate for entertainment. *While image degradation is often desired for the illusory effects necessary to suspend belief in the telling of a story, for scientific, medical, and many industrial and military applications the need for true, undistorted images is critical.*

In all of the presently available video systems, the image capturing device (the camera), the transmission technique and the display appliance are locked together in synchronism: the image is captured by scanning a light-sensitive surface on which the image has been projected; and in the display (the television set) a similar, exactly synchronized beam of electrons excites the cathode ray tube's phosphors. Any amplitude distortions, temporal delays (e.g. electromagnetic wave reflections from nearby buildings) or other defects occurring at any point in the chain between camera and display immediately show up in the picture.

Variable aspect ratios are fundamentally impossible because the synchronized scan rate has to conform to fixed display parameters.

A further characteristic of conventional television is the use of *interlaced scanning*. Interlacing was required in the earliest days of crude television to provide the bandwidth compression necessary to transmit the video signal. Interlacing, basically an obsolete form of compression, creates a picture that is especially sensitive to phase and amplitude distortions as can be readily seen in the moiré and fringe artifacts often present in television pictures.

Moreover, *interlaced images are missing half of the full image at any discrete point in time.* This makes lossless compression essentially impossible, and also adds visual artifacts created by missing information improperly processed by the human eye/brain nexus. This usually shows up in various forms of flicker and dot crawl that can range from mere annoying to the inability to see critical events on the display. Even on very high-resolution TV systems, at critical spatial frequencies interlace will produce fuzzy edges on rapidly moving objects. So, with

interlace, it is difficult or impossible to align computer-generated, progressively scanned overlays on real life images without generating visual vibrations and dot crawl. These most annoying artifacts are exacerbated by the non-square imaging rasters used by all other video systems except ours.

Summary

In summary, close synchronism means that optimum signal or image processing techniques for each element of the chain are difficult to apply. *Synchronism, coupled with interlace is especially a problem for image compression processes.*

All current television or video systems, including all the commercial broadcast HDTV proposals (save one) use some variation of these synchronous and interlaced scanning mechanisms. The one set of HDTV standards permitted by the FCC which does *not* use interlace or a non-square raster is the set followed in our cameras. So far, after one year of demonstrations, ours are the only such devices in existence.

With modern digital transmission, synchronism, interlace and fixed aspect ratios are no longer necessary, but historical forces have kept current video systems and most proposed HDTV systems from taking advantage of completely decoupled architectures.

The requirements of modern scientific imaging demand high-performance cameras without introducing artifacts and false images. Our systems' wide range of sensitivity, resolution, and color balance, coupled with its inherent digital design friendly to computer processing and appliances, will enable new markets based on image enhancing. A progressive-capture, high-resolution camera opens the pathway to vastly improved performance and productivity for existing and new applications.

Clark E. Johnson; 612-922-8541; fax 612-922-8820; e-mail clark@rpcp.mit.edu
Richard Jay Solomon; 617-253-5159; fax 413-267-5172; e-mail rjs@rpcp.mit.edu

DESIGN AND IMPLEMENTATION OF A 3-CCD, STATE OF THE ART, 750-LINE HDTV PROGRESSIVE SCAN BROADCAST CAMERA

S. M. Spitzer, J. Toker - Polaroid Corp, Cambridge MA
A. Moelands, P. Centen, J. van Rooy - Broadcast Television Systems,
Breda, The Netherlands
B. J. Gerovac - MIT, Cambridge MA

ABSTRACT

This paper describes an HDTV camera system implementing the 1280 x 720 image format supporting the 750/60/1:1 production standard. An interlaced HDTV camera has been adapted to meet the proposed US HDTV progressive standard. A new 1" video format, 16:9 aspect ratio progressive scan, frame-transfer (FT) CCD sensor with square pixels was designed, and sensor incorporation and camera adaptations were implemented. The new sensor is described and the impact of the 750-line standard is discussed. The first prototype of this camera will be demonstrated at the 1996 NAB show.

INTRODUCTION

The CCD image sensor was introduced into broadcast cameras around 1986. In 1992 the first HDTV CCD camera system to meet the proposed (European) EUREKA HDTV standard was introduced^[1]. This LDK 9000 camera system, designed by Broadcast Television Systems (BTS), is based on a 1" frame-transfer sensor that was developed to support the 1250/50/2:1 interlaced European standard as well as the new American 1920 x 1080 format 1125/60/2:1 interlaced production standard. This camera has been used for the production of test material by the Advanced Television Test Center.

The Advisory Committee for Advanced Television Services (ACATS) recently released its proposal for a new U. S. television standard commonly referred to as HDTV^[2]. The development of a progressive scan format for HDTV has been driven by concerns such as compatibility with computer systems, ease of compression, freedom from flicker (especially with graphics), and better temporal resolution. As discussed in the ACATS process, the interoperable and extensible HDTV system can serve not only entertainment and television, but can also offer economic and qualitative benefits to education, health care and human services, commercial enterprise, and the information infrastructure.

Several factors were considered critical to achieving interoperability^[3]. One of these factors is the use of progressive

scan square pixel image formats in capture, transmission, and display. Thereby, the television equipment can be extended to and stimulated by applications in computer communications, high quality imaging, synthetic imaging, animation, motion pictures, and so forth. The information infrastructure needs an image architecture that eases exchange between industries and applications.

Significant technical hurdles have acted as barriers to deploying a progressive scan HDTV system. Nonetheless, the Grand Alliance did incorporate progressive scan among their formats. Most experts agree that a progressive scan system is ultimately desirable and certainly inevitable in the proposed lifetime of HDTV, though the time frame is debated.

The major technical hurdle has been the difficulty in producing a progressive scan camera of comparable sensitivity and specifications to a studio quality interlace scan camera. Existing commercial and prototype cameras have been inadequate. Indeed, the existence of adequate component technology has been in doubt. Herein laid the motivation for our research and development efforts.

Through a cooperative effort between the Polaroid Image Sensor Technology Division and BTS the LDK 9000 HDTV CCD Camera system was recently adapted for the progressive 1280 x 720 standard (750/60/1:1). A new frame-transfer sensor meeting this standard was developed at Polaroid to be optically, electronically, and mechanically compatible with the previous interlaced sensor, although differing in image format and timing. The main adaptations of the camera system performed by BTS included the following elements:

- Camera/sensor pulse generator
- Camera Processing Unit pulse generation
- Vertical contour delay
- 7 inch view finder

Apart from the above-mentioned functions, minor adaptations were made in several areas to meet the timing specification and to optimize sensor performance.

Parameter	Value
Aspect Ratio	16:9
Interlace	1:1 (progressive)
Field frequency	60 Hz
Total number of lines	750
Number of active lines	720
Line frequency	45,000 Hz
Total line time	22.222 μ sec (1650 samples)
Active line time	17.239 μ sec (1280 samples)
Horizontal blanking	4.983 μ sec (370 samples)
Sample frequency	74.25 MHz
Sync pulse	Tri-level

Table 1. Main characteristics of the 750/60/1:1 1280 x 720 production standard

PROPOSED 1280 X 720 PROGRESSIVE TELEVISION STANDARD

Interlace scanning has proven to be an efficient way of sampling pictures. The flicker perception of the human eye demands a refresh rate of the CRT of at least 50 times per second to prevent large area flicker. In order to save bandwidth, it was decided to refresh alternately the odd lines and even lines, thus doubling the vertical resolution for a given signal bandwidth. This means that for a given signal bandwidth the number of pixels in an interlace standard will be twice the number of pixels in a progressive system, resulting in a better static resolution.

But interlace scanning also shows some well-known artifacts, especially with moving pictures:

- It is impossible to combine two fields to one picture for moving objects, as each field comes from a different moment in time. This is a major drawback for creating still pictures from a moving scene, and for video to film transfer.
- While, with proper filtering, the frame can be nearly free of aliasing, each field may contain aliasing since it has only half of the samples in the vertical direction. The human eye has to integrate out aliasing effects per field, to see the full frame resolution of the picture. This results in small area flickering at field rate. The canceling of aliasing between fields only holds for still pictures.
- Even at slow vertical movement of one line per field vertical aliasing is dramatically increased. This is especially visible on slowly moving almost horizontal lines in the picture.

With the move to digital television, the performance of the compression system becomes critical. Compression of interlaced signals is more complex and performs worse than compression of progressive scanned signals, where the entire image is sampled at the same time.

The 750-line 1280 x 720 format progressive scan standard provides a good, practical solution to the problems of interlacing while obtaining excellent compatibility with interlaced HDTV^[4]. The key characteristics of this standard are given in Table 1^[5].

Several features of this standard greatly add to its practicality. Firstly, the picture format uses exactly 2/3's the horizontal and vertical pixel counts of the interlaced standard for ease in resampling. Next, the field and pixel frequencies are identical allowing the use of the same production equipment. In addition the line time is 3/4's that of interlaced potentially easing analog delay designs. Finally, adequate horizontal and vertical retrace intervals are allowed.

CCD SENSOR

The CCD sensor was designed specifically for progressive scan high definition video applications. With square pixels, an active array of 1280 x 720 pixels (1296 x 730 total pixels), a 16 mm diagonal for use with 1" format lenses, and with 60 frames/second operating speed, this sensor is ideal for the proposed 750/60/1:1 progressive-scan HDTV standard (Table 2). The frame-transfer architecture used provides high sensitivity, high fill-factor, no lag, and no smear when used with a mechanical shutter wheel as in the LDK 9000. The imager (shown in Figure 1) consists of the imaging array with both active and dark reference pixels, a full resolution storage section, a dual-channel horizontal register, and two output buffers.

The imaging pixel (Figure 2) is a 3-phase buried channel

Parameter	Value
CCD-type	FT
Optical format	1 inch
Image diagonal	16 mm
Image area width	14.00 mm
Image area height	7.88 mm
Number of lines	730
Pixels/line	1296
Pixel width	10.8 μ m
Pixel height	10.8 μ m
Chip width	15.29 mm
Chip height	15.25 mm
Chip area	233 sq. mm
Output registers	2
Pixel output rate	74.25 MHz
Frequency H-clocks	37.125 MHz
Swing H-clock	5 V
Frequency V-clocks	2.475 MHz
Swing V-clock	10 V

Table 2. CCD characteristics

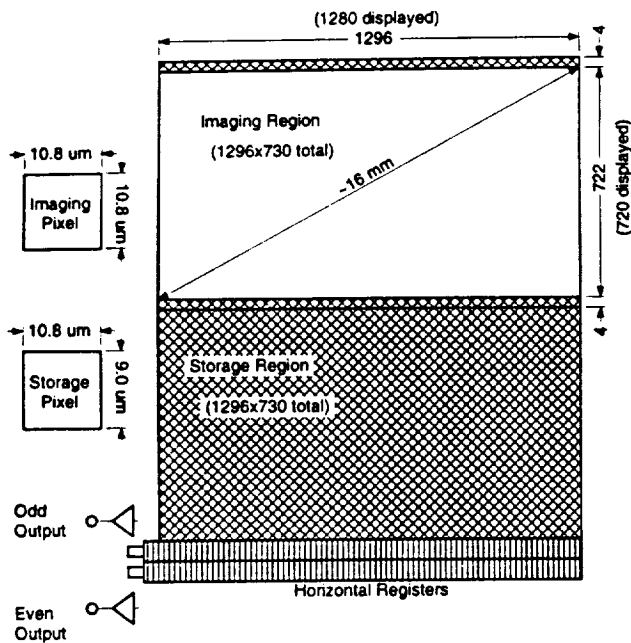


Figure 1. CCD Block Diagram

device with integrated vertical anti-blooming protection. It is 10.8 x 10.8 microns square. The device is formed with three polysilicon layers, one for each phase, with large open areas (>33% of pixel) for enhanced blue light sensitivity. The vertical N-type buried channels are separated by P+ channel stops. The P-Well doping is modulated to form a weak spot in the center of the channel that acts as the anti-blooming barrier, which turns on when the pixel fills up to drain excess photocurrent down into the lightly N-doped epi layer. The storage pixel is configured similarly, although it uses wider poly gates for greater charge storage density. Thus the storage pixel could be made smaller (9.0x10.8 microns).

Since this sensor was designed to be compatible with the already existing camera, the process was carefully adjusted to give proper operation at the supplied clock voltages. This was complicated by the large number of functions that the pixel must implement: light absorption, charge collection, vertical overflow drain, charge transport, and charge reset (frame clear) for exposure control.

High vertical transport shift frequency (2.475 MHz) is required to move the charge from image to storage section during the brief optical blanking period provided by the shutter wheel. This frequency is by necessity higher than that used in the interlaced design because there are more lines to move (730/frame versus 589/field). Two-level aluminum wiring was used to shunt the polysilicon gate resistances resulting in less than 1 ohm equivalent series resistance. This allows the roughly 6 nF capacitive load to be driven at the required speed. Narrow aluminum straps connecting the poly gates run over the channel stops in the imaging section so that they have minimal impact on light

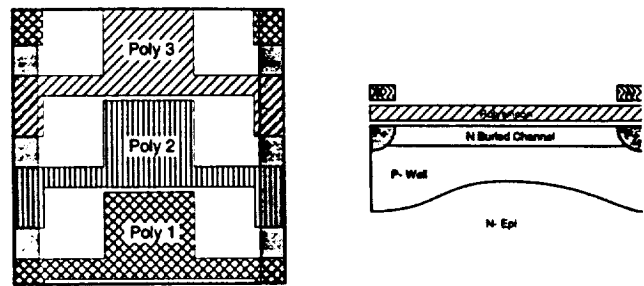


Figure 2. Imaging pixel top view (left) and cross-section (right).

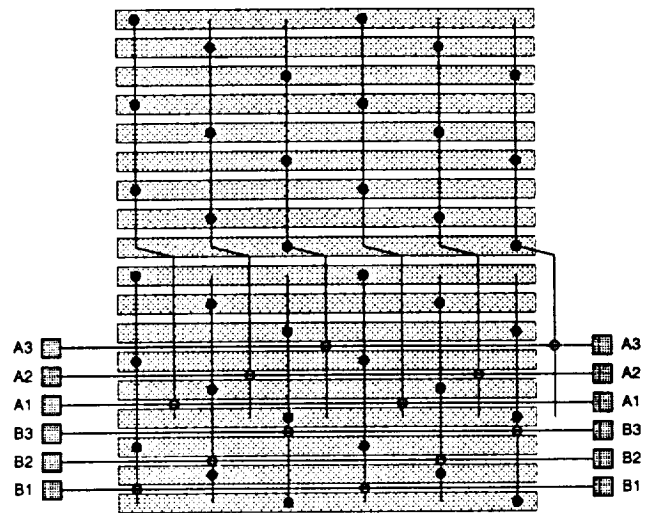


Figure 3. Double-level metal clock interconnect

sensitivity. Care was taken in their design to ensure that no fixed pattern artifact such as stripes was introduced by the straps. These are connected by busses on the second aluminum layer running across the storage section as shown in Figure 3. These busses are tied to package leads at each end to further lower series resistance.

During readout the charge is shifted one line at a time from the storage section into the horizontal registers. These charge packets are split up into the two registers on an even-odd column basis, using just a single transfer gate. The two registers operate in parallel at one-half the pixel frequency (37.125 MHz). The charge is transported along the 4-phase buried channel register to the matching sense nodes. The camera provides 4 phase clocking with 4-5 volt amplitudes, which posed a greater challenge in this design because the square pixels are wider and hence the horizontal gate length is greater reducing the fringing fields which assist charge movement. Extensive simulation was performed to ensure that excellent horizontal charge transfer would be achieved even at high clock frequencies.

Parameter	Value
Sensitivity (at sense node)	14 uv/e-
Amplifier gain	0.4
Noise after DLP in 30 MHz	33 e-
Bandwidth output	150 MHz
Quantum efficiency (peak)	26%
Sensitivity with BG40	2150 e-/lux
Overexposure	100,000 X
Full well capacity	40K e-
Dynamic range	62 db
Sampling frequency vertical	92.6 line-pairs/mm
Sampling frequency horiz.	92.6 line-pairs/mm
Image lag	none
Smear (incl. camera)	none

Table 3. CCD Performance

The output buffers are fairly conventional three-stage source-follower design. All three drive transistors are surface channel, giving high transconductance for low noise operation. The bandwidth (>120 MHz) is high enough to ensure accurate signal transmission. The layout of the two buffers was arranged to ensure they would match even with layer-to-layer misalignment during fabrication. The two video output signals are combined in the video pre-processor using the delay line principle (DLP).

Measured performance of the initial samples of the CCD sensor is summarized in Table 3.

OPTO-MECHANICAL DESIGN

The opto-mechanical system of the 1" CCD HDTV camera is designed to use lenses with a maximum aperture of f/1.2. The system consists of (from front to back): seal glass, IR-filter, retardation plate, shutter wheel, two 4 position filter wheels (for effect and ND filters), beam-splitter, optical low pass filters and sensors. In the adaptation of the optical system to the progressive format, in addition to the sensors only the optical low pass filter was changed.

Modulation Transfer Function

The modulation transfer function (MTF) of the camera is determined by the lens, optical low-pass filter, aperture of the image cell, and the electrical sample-and-hold. The MTF of the lens at f/4 is mainly diffraction limited. The MTF of the optical low pass filter is cosine shaped. The aperture of the image cell and the sample and hold both have $\sin(x)/x$ characteristics. Based on this model one expects a MTF of 47% for a sine wave at 27 MHz, versus a measured value of 50% (Figure 4).

Aliasing

A CCD-camera is a two-dimensional spatial sampler. The kings of fashion do not care about Nyquist nor does Nature! Therefore in everyday life the Nyquist condition - that the

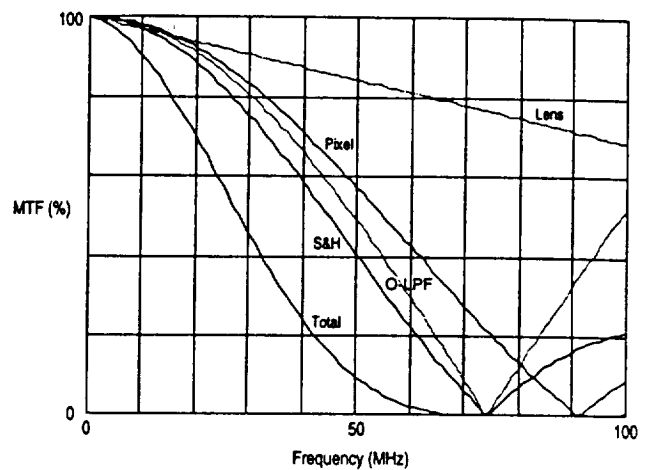


Figure 4. The modulation transfer function (MTF) of the camera. Shown are the separate contributions of lens, optical low-pass filter (O-LPF), aperture of the pixel, and the sample-and-hold.

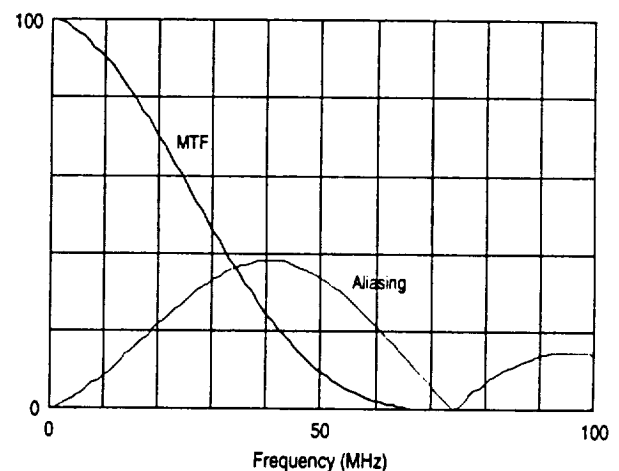


Figure 5. Shown are the MTF of the camera-head and the residual aliasing due to folds at the pixel sample frequency of 74.25 MHz

maximum frequency of the optical signal must be below half the sampling frequency - will be violated. This will cause Moire, or aliasing, patterns, which will create low-frequency patterns the eye is very sensitive to.

The frame-transfer image cell has a large aperture and therefore has intrinsically good horizontal and vertical aliasing behavior for higher spatial frequencies (greater than the Nyquist frequency). An optical low-pass filter helps to reduce aliasing further, especially at lower frequencies, by introducing dips (or notches) in the MTF. These dips must be at the vertical sampling frequency (92.6 line-pairs/mm) and at the horizontal sampling frequency (74.25 MHz, or 92.6 l-p/mm) for maximum effect (Figure 5). The need for a vertical anti-alias filter is unique to this progressive scan camera, since interlaced CCD sensors typically have considerable overlap in the even and odd scanning apertures that performs a similar function at the cost of vertical resolution.

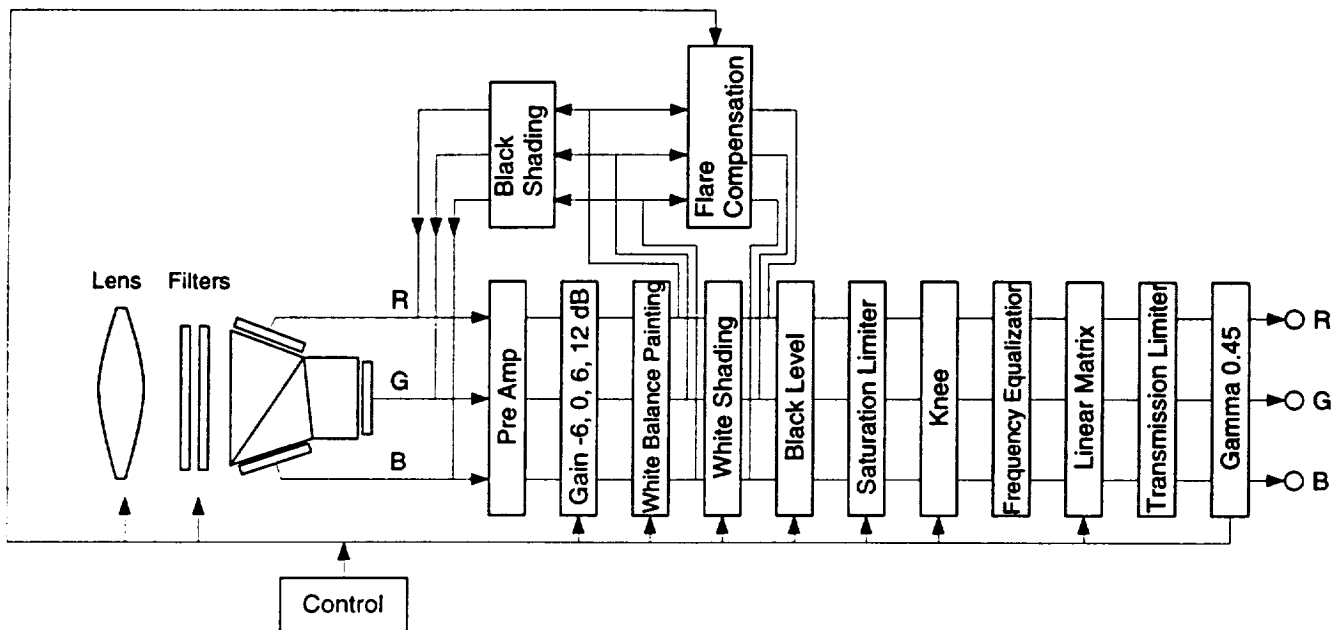


Figure 6. Camera head video processing.

VIDEO PROCESSING INCLUDING CONTOURS

Video processing in a progressive scan camera is not very different from the processing in an interlaced camera. The processing of the LDK 9000 camera has already been described in an earlier paper [1]. It consists of a part in the camera head (Figure 6), and further processing in the camera processing unit (CPU) (Figure 7).

Important design objectives for the LDK 9000 video processing were:

- Gain control over full temperature range.
- High dynamic range.
- Headroom before highlight compression of more than 14 dB.
- Signal/Noise deterioration due to video processing less than 1 dB.
- High quality, reliability and operational flexibility.
- Low power consumption.

Operation following the 1280 x 720 progressive standard calls for some specific adaptations as compared to the 1920 x 1080 interlaced standard:

- Line time is changed from 29.6 usec. to 22.2 usec. This calls for different line delays in the contour delay unit.
- Active line time is changed from 25.8 usec. to 17.2 usec. This calls for a more accurate timing in the video processing as timing errors will be more visible on the display.

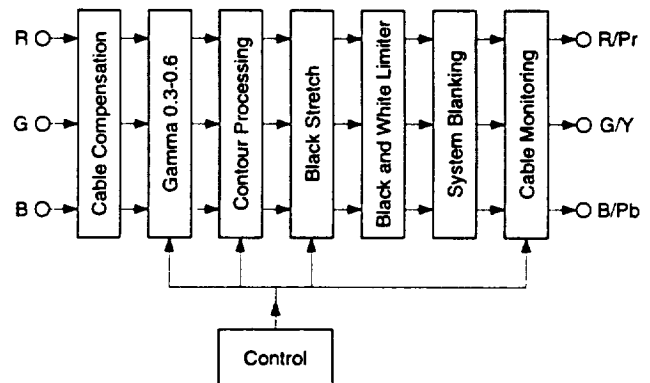


Figure 7. Video Processing CPU

- Vertical contours will look different -- the vertical contour generation in a 1080 line system is field based, with the 0T, 1T and 2T lines 1/540 picture height apart. In a 720 line progressive system vertical contours are generated from lines with a spacing of 1/720 picture height. This results in a higher vertical peaking frequency for vertical contours in a 720 progressive system.
- Horizontal contours will have a lower spatial frequency peak in the 1280 x 720 progressive scan system. This can be changed by shortening the delay lines in the contour processor, but there are practical limitations imposed by the lower Nyquist frequency of the 1280 x 720 system.

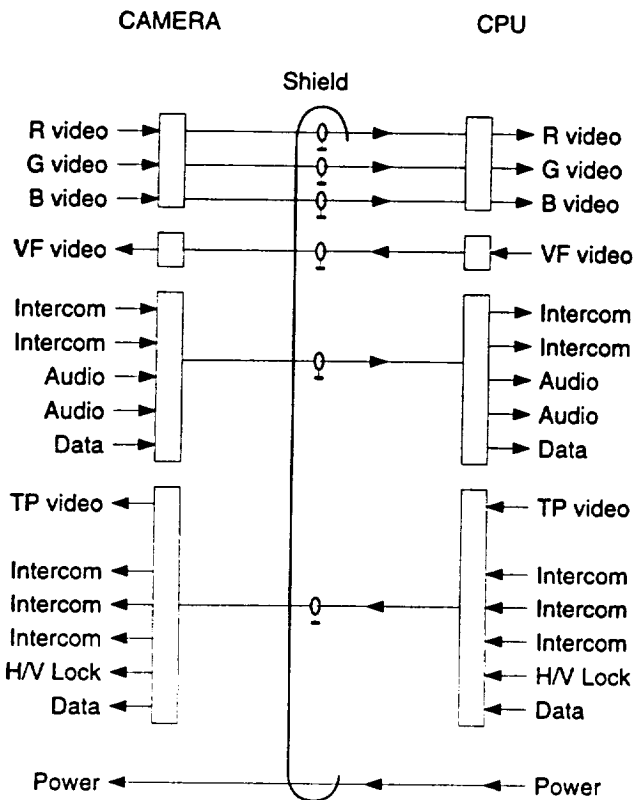


Figure 8. Multicore Transmission system

SIGNAL TRANSMISSION

The connection between camera and processing unit is formed by an interconnection system of cable and electronics specifically designed to maintain signal quality. A multicore cable can be used for short distances up to 300 meters. For longer distances, the multicore cable can be extended with a fiber optic system. Four coaxial cables are required for the R, G, B and view finder video signals. The remaining signals (figure 8) could be multiplexed into a single bidirectional coaxial cable, or into two separate single directional cables. This latter approach is used to achieve a simple interface with an optical fiber. Additionally, power wires are added, yielding a custom-made circular cross-section multicore cable.

The electronics provides automatic compensation for all multicore cable lengths between 0 and 300 meters. This is realized by dividing the total compensation into a fixed part and an adaptive part. The fixed part can compensate any cable length within an increment of 12.5 m. The compensation determined at power-up, by means of a successive approximation measurement. The adaptive part, which is independent in each channel and continuously active, has two functions:

- It has to compensate the last residual cable length within the resolution of the fixed part,

- It has to compensate (the frequency dependant) loss differences as caused by such things as temperature changes of the multicore cable and differences between the individual coaxial lines.

Delay differences between the coaxial lines is kept small by using high quality coax: maximum 1.5 ns between R, G, and B video signals at 300 meter cable length.

7 INCH VIEW FINDER

The main challenge in adapting the view finder to the progressive scan 1280 x 720 format was operating at much higher line frequency (45 KHz) given limitations on power dissipation and demands for high brightness and contrast. A stable high voltage source is required to prevent "breathing" at high beam currents and to secure high resolution performance.

Although spatial frequencies are lower for 1280 x 720 than for the 1920 x 1080 system (27 MHz bandwidth gives 780 TVL for 1920 x 1080 versus 520 TVL for 1280 x 720), focus assist is still a valuable tool for the camera operator. Apart from peaking in the view finder, the HDTV camera system is provided with two focus assisting tools:

- **Magnifier:** Momentary activation of this function enlarges the center part of the image by approximately 1.6 times, filling the whole screen.
- **Crawler:** Small details in the picture are converted to a more coarse structure, which gives edges and other fine details a highly visible crawling pattern. Optimum focus is obtained when this crawling serration reaches the maximum intensity. It acts more or less like "peaking" and can be used continuously.

CAMERA SYSTEM

The camera being presented is part of a complete system configured for broadcast applications. The system, as modified for the progressive standard, consists of the following system components:

- Camera head
- 7-inch view finder
- Camera Processing Unit (CPU)
- Multicore cable
- Master Control Panel (MCP)
- Operational Control Panel (OCP)
- Lens
- Accessories

The camera head has been designed as a compact, lightweight, modular unit (Figure 9). The camera features two four-position filter wheels with three neutral density filters and two special effect filters.

The 7-inch view finder can be mounted on a optional specially-designed support above the camera. The support is designed to accept the lightweight camera combination. The camera, with an optional 1.5-inch view finder, can be

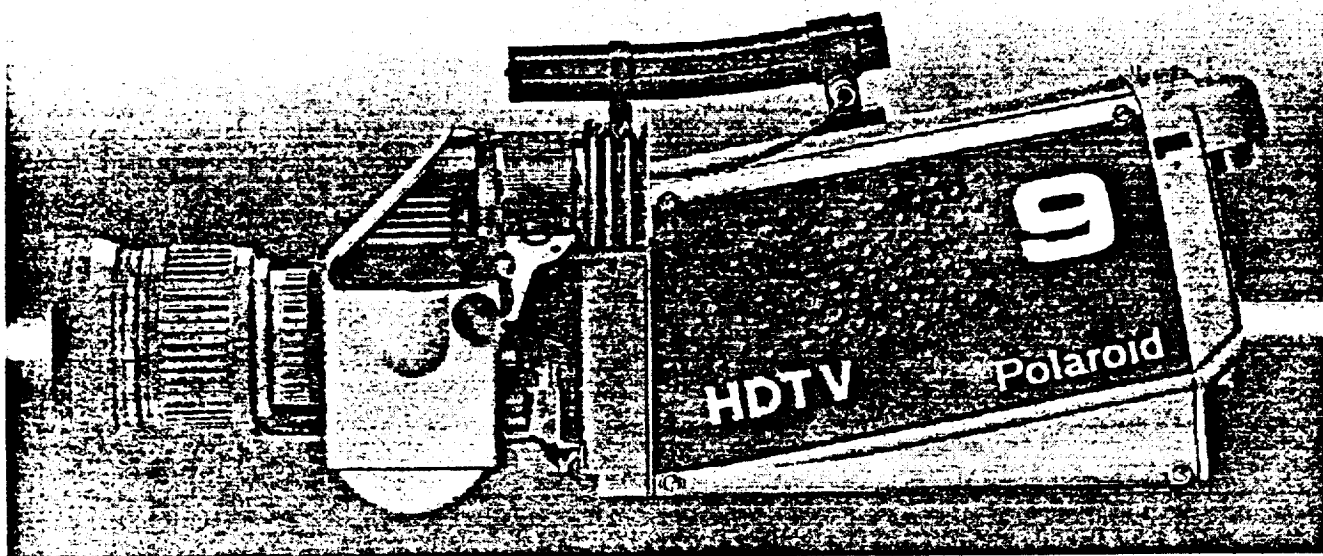


Figure 9 Camera head

easily placed upon or taken out of this support, leaving the support and 7-inch view finder on the tripod.

The CPU is a 19 inch rack mountable component which is 3 standard units high. The device is constructed using standard Eurocard PC boards and a rear connector panel with all signal interconnection options commonly used in broadcast studios.

The control panels follow the Series 9000 control philosophy, as used with all BTS standard TV cameras. The Master Control Panel gives access to most of the set-up controls via menus. The Operational Control Panel (OCP) provides all the operational control functions of the HDTV camera. The OCP is arranged with user-friendly directly accessible controls.

The camera can be used with a wide range of lenses built with internationally standardized interfaces. The camera presented is equipped with an 11 x 11 barrel-type lens from Fujinon. This heavier barrel-type zoom lens is supported by standard film-style accessories: a bridge plate underneath the camera accepts support rods, lens supports, matte boxes, etc.

The main camera system characteristics are summarized in Table 4. Performance specifications are summarized in Table 5.

For picture evaluation during the development period a Barco color monitor, Reference Calibrator model 121, was used. This monitor is capable of displaying 1280 x 720 progressive signals without effecting picture quality. Noise measurements were done on the Rohde & Schwarz VNA (Video Noise Analyzer). The Tektronix 1730HD Waveform monitor was used.

- An electronic white balance range from 2500K to 15000K.
- Highlight compression in automatic and manual mode.
- Black stretch in Y and R,G,B.
- Colorimetry according to EUREKA/EBU standard.
- 2-Dimensional contours.
- Electronic shutter with 5 and 2 msec exposure time. Also 50 Hz and 60 Hz lighting positions are available.
- Camera power consumption approximately 22 W.

Table 4. Main camera system characteristics

- Modulation Transfer Function of over 40% at 520 TVL (27 MHz) without contours.
- Limiting horizontal resolution of 700 TVL.
- Sensitivity of 1200 Lux at F/4.
- S/N ratio of 50 dB at a bandwidth of 30 MHz.
- The max. lens aperture is F/1.2.
- Dimensions approx. 140 x 210 x 350 mm.

Table 5. Camera system specifications

CONCLUSION

We report here on the first CCD HDTV broadcast camera to demonstrate the newly recommended 1280 x 720 progressive standard (750/60/1:1). A frame-transfer CCD was custom-designed, and a broadcast-quality interlaced HDTV camera was modified to meet the progressive standard.

A prototype camera has been built and is demonstrated at the 1996 NAB Exhibition. This camera meets all specifications as presented in this paper, and meets all goals toward proving the feasibility of the 750-line progressive HDTV standard.

The LDK 9000 system, with the 1250/50/2:1 standard, has already been in use for 4 years in Europe. During this time these systems have been used, to complete satisfaction, at a wide variety of events. A modified system for the new American 1920 x 1080 production standard (1125/60/2:1) was used for the production of test material by the Advanced Television Test Center. It is anticipated that this proven record of reliability and success will carry over to the progressive scan camera.

This project represents a successful embodiment of collaboration between industry, university, and research laboratory to accomplish more in a shorter period of time than any one could do alone.

ACKNOWLEDGMENTS

The authors acknowledge a number of people for their valued contributions to this project. They include: Selim Bencuya, Mary Finn, Richard Hoyer, Chris Needham and William Vetterling of Polaroid; Theo Kedziersky, Flip Stok, and Chris Taylor of BTS; and Jae Lim of MIT.

This project was supported in part by the Advanced Research Projects Agency.

REFERENCES

- [1] A high performance full bandwidth HDTV camera applying the first 2.2 million pixel FT-CCD sensor., J. Blankevoort, H. Blom, P. Brouwer, P. Centen, B. vd Herik, R. Koppe, A. Moelands, J. v. Rooy, F. Stok, A. Theuwissen. Presented at the 134th SMPTE Technical Conference, November 1992, Toronto Canada.
- [2] ATSC Digital Television Standard, September 16, 1995
- [3] Advisory Committee Final Report and Recommendations, ACATS, November 28, 1995
- [4] 1280 x 720 Progressive: A Reevaluation, W. Miller, SMPTE Journal, October 1995
- [5] Proposed Standard for Television - 1280 x 720 Scanning and Interface, second draft SMPTE document S17.392-1147B, October 4, 1995.

MIT/POLARIOD DIGITAL IMAGING SYSTEM

The MIT/Polaroid camera that you have seen demonstrated is the culmination of four years of work on a complete digital imaging chain. Supported by MIT and funded by ARPA, this new approach to image capture and display consists of the following components. While only the camera has been fully implemented, the other essential elements have all been designed. The underlying criteria was to develop a fully digital image capture, storage and display system with a wide range of capabilities with capabilities that are user and application selectable.

Motivation for this work partially arose from the extreme difficulties experienced by development of an HDTV standard that met the necessary requirements of extensibility, scalability and interoperability. Only by fulfilling these essential criteria can future imaging technologies be painlessly introduced into the system.

The imaging system consists of four components.

1. Image Capture. The MIT/Polaroid camera consists of an advanced CCD image capture array. Three of these 720 x 1280 (effective) pixel CCD's are carefully mounted on a beam splitter and each is preceded by a dichroic filter to select the three primary colors. An electronic shutter whose speed can be adjusted from infinity to 1/200 of a second permits the camera to be used as a still camera or as an electronic movie camera.

A significant advantage of this camera over conventional TV studio cameras is its frame capture capability. That is, it operates like a high-speed "still" camera. All of the image is captured at once; in contradistinction to a TV camera where the image is scanned, point by point, during the 1/30 second frame time. The image scanning of current conventional television systems is an absolute requirement resulting from the need to synchronize all steps of the process from image capture, through transmission, display and, if required, storage on a VCR.

This digital imaging system completely decouples the various components. No synchronization is required among any of the system components. Thus this imaging system meets the requirements of interoperability, scalability and extensibility.

2. Display. The MIT/Polaroid camera demonstration uses a high-resolution color computer monitor. There are a wide variety of other displays available including the Texas Instruments DMD (digital mirror device) and the Hughes/JVC light-valve projector. Display technology is currently adequate to provide full rendition of the camera's capability.

3. Storage. Most applications require preservation of the image for display/analysis at a later time. Currently there is no commercially available data recorder that will handle the full-resolution data rate of the MIT/Polaroid camera operating at 60 frames/second (1.2Gb/second) directly. The Panasonic D-5 recorder can be altered to work, requiring some electronic modifications along with a format converter.

4. Image Compression. Some degree of compression can be used in all imaging applications. "Lossless" compression (that with no loss of information) yields about a 4 times data reduction. Further compression has some effect on image quality, however, by taking advantage of the human eyes' sensitivity variation with position and motion, compressions up to a factor of 100 or more can be used without visible degradation of the image. The use of compression is very application-dependent. Presently, a 200:1 software-implemented compression scheme can be demonstrated. Its conversion to silicon (hardware) is an essential component of a future development effort.

5. Transmission. There are many methods to get the image signal from the camera to the recorder and display, ranging from fiber cable to broadcast to coaxial cable. In general, the

transmission method determines the bandwidth available. However, since the system is scalable, a bandwidth-limited transmission simply reduces the quality of the displayed image.

SOME APPLICATIONS

A. "Realpresence" Videoconferencing.

This application links the MIT/Polaroid camera to a projection display such as the BARCO high-resolution projection system. Incorporated into properly designed conference rooms with "surround-sound," it would provide outstanding teleconferencing capability. The system would be designed to minimize bandwidth consumption by using variable frame rate image capture (no sense in using high frame rates when everyone is snoozing!) along with variable compression.

Such a high-resolution video-conferencing system would have application in a wide range of customer environments. As an example, the textile and garment industry might use video-conferencing to cut down turn-around time between nascent and rapidly-changing fashion demands. One can easily envision designers scattered throughout the world connected directly to the factories in North Carolina or the Far East.

B. Remote Education, Professional Training and Remote Medical Diagnosis

Many studies have shown that high-quality images, approaching reality, can enhance and may actually be essential to the use of computer-based teaching tools. Current NTSC and the proposed HDTV television systems cannot fulfill the image quality requirement.

Remote medical diagnoses is an essential component of a nation-wide health-care program. It is essential that the same quality of medical care be available to rural areas as to residents of large cities. With the continuing refinement of medical technology, much of which is image-based (e.g. MR scans, X-rays, ultrasound—to name only a few), a real-time, bi-directional, high-resolution image capture, transmission and display system is essential. The imaging system described herein has all the components necessary to satisfy the medical imaging requirements.

C. Museum and Historical Artifact Archiving

The best approach for developing the techniques and technology for museum archiving is to use a high-resolution imaging system. By taking a "picture" of the object (or, for that matter a building) from several angles and then post-processing the various digital images, a true three-dimensional image is generated. All images are stored for later use. The viewer then has the ability to "move" the image around and "see" it from any angle.

D. Military Reconnaissance

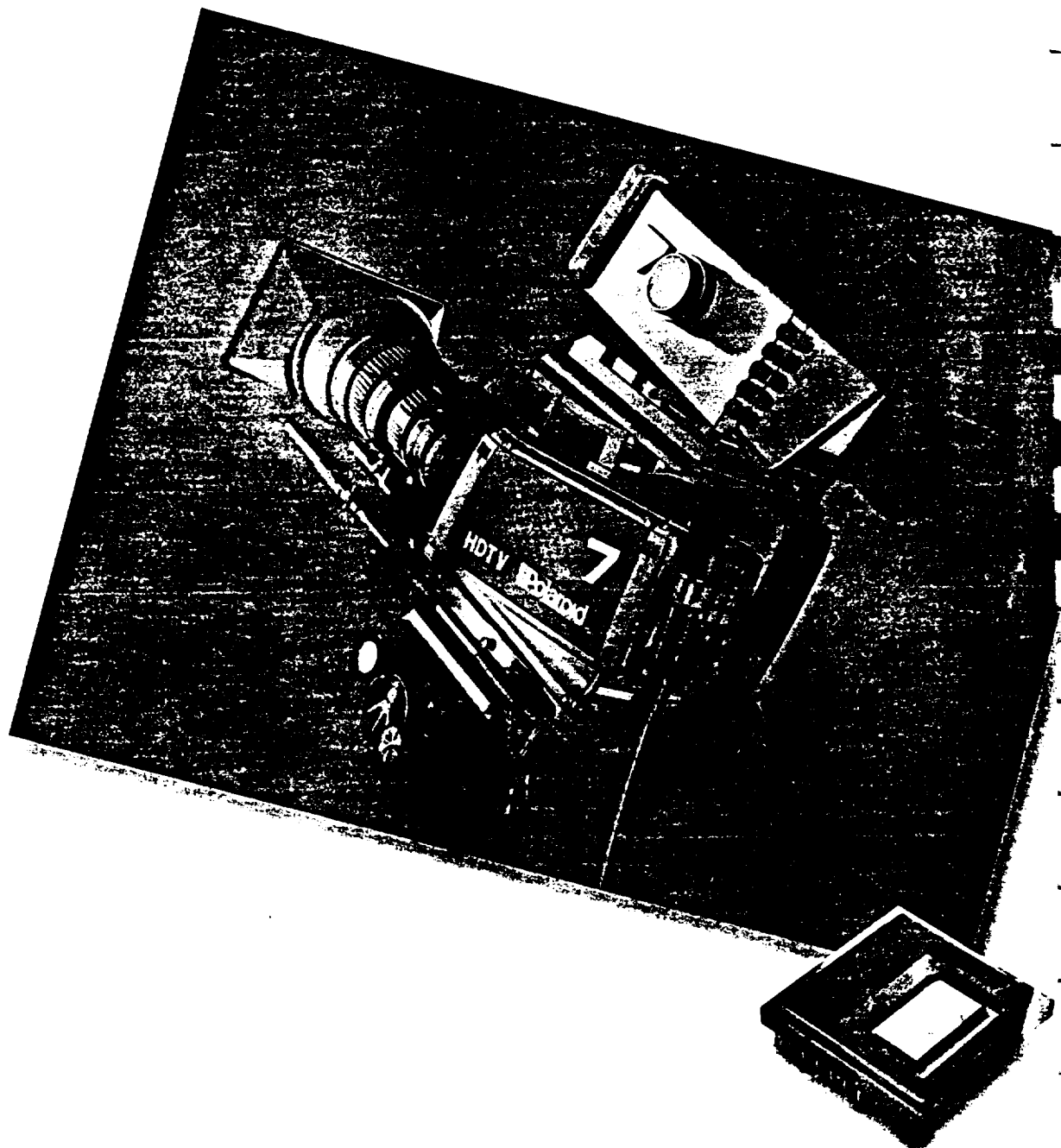
In military surveillance operations, size, weight and light sensitivity are the most important parameters. Recent developments in CCD sensor technology would provide the same sensitivity on the CCD's receptive area about 25% of that used in the present camera implementation. This reduction in size permits much smaller lenses as well as a much reduced camera body. The decrease in signal-to-noise ratio inherent in the smaller CCD pixel size can be overcome with improved signal-processing techniques.

E. Other Applications

For the first time, it is now possible to have both a "still" camera and a "movie" camera in the same body. The tradeoffs between exposure time and sensitivity that we are all familiar with in 35mm cameras exists here. Only now the images can be enhanced electronically and the color "gamma" adjusted to the user's requirements. A single-body, dual-use movie/still camera has many applications. Those of the military, where space and weight limitations are critical, is only one example.

Just when you
thought you
had seen it
all...

PTC-9000 offers a level of performance and image quality unavailable today. This camera system meets the newly approved 750 line progressive scan High Definition Television (HDTV) standard providing a 1280 x 720 resolution image, in a 16:9 format, at a rate of 60 frames per second.



PTC-9000 Specifications

Camera System

Production standard

- 750 line/60 frames/sec/1:1 progressive

Image Format

- 1280 x 720 square pixel

Aspect Ratio

- 16:9

Optical System

- 1-inch format
- Image diagonal 16mm
- Standardized lens interface acc. EBU, BTA, SMPTE
- Prime lenses, barrel and box type zoom lenses up to f/1.2
- Low optical axis, approx. 90mm
- Accepts Arri film style accessories.
- Two motor-driven filter wheels, each with 4 positions
- Optical low-pass filtering

Video Performance

Sensitivity

- Approx. 1200 lux at f/4

Maximum Sensitivity

- 25 lux at f/1.2 + 12 dB gain

Signal to Noise

(at normal gain)

- Approx. 50 dB in Y

Modulation Depth

- 40% at 27 MHz

Registration

(without lens errors)

- Typical 0.05% of picture height

Exposure Control

- Nominal, 50 Hz, 60 Hz, 1/200 s, 1/500 s

Video Signals

Camera head output

- Viewfinder video (return video or local green)
- Monitor/Viewfinder output
- Teleprompter 5 MHz (with large lens adapter)

Camera Processing Unit

- Output
R,G,B,S
Y, Cr, Cb
Switchable R, G, B / Y, Cr, Cb
Picture Monitor
Waveform Monitor
- Input
Ext. 1, 30 MHz bandwidth
Ext. 2, Teleprompter, 5 MHz bandwidth
Reference

Audio

(Available only with Large Lens Adapter)

Camera to CPU

- Two high quality audio channels with remote gain control

Phantom power

- 12 or 48 V DC

Intercom

CPU to Camera Head

- Engineering
- Production
- Program

Camera Head to CPU

- Camera operator
- Floor manager (large lens adapter)

Maximum Cable Lengths

Multicore

- 985 ft. (300 meters) camera to CPU

Fiber Optic System

- Max. 6,560 ft. (2 km) with remote power, 4x1,640 ft. (4x500 meters)

Viewfinders

- 1 1/2" and 7", with focus assistance: crawler, magnifier

Power supply

Line Voltage

- 115-230 V+/-15%, 47-63 Hz

Power Consumption

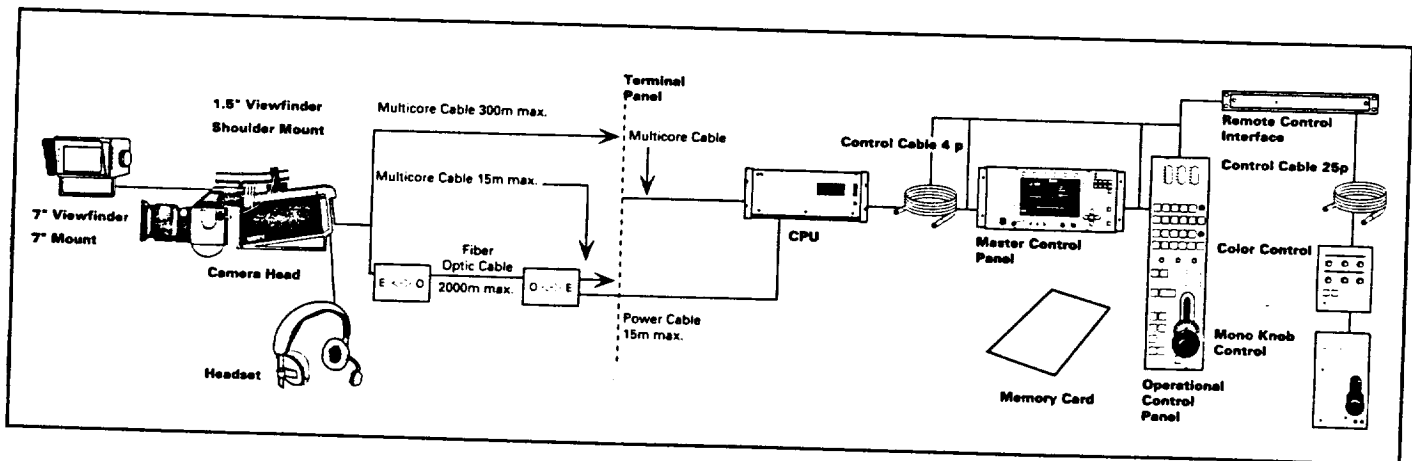
- CPU: 100-300 watts, depending on camera configuration
- OCP: 5.5 watts
- MCP: 23 watts

Approximate weights

- Camera Head: 15 lbs.
- CPU: 33 lbs.
- OCP: 33 lbs.
- MCP: 13 lbs.

These typical specifications are subject to change without notice.

For further information please call us at (617) 386-8747 or by fax (617) 386-3900



Polaroid Corporation, Image Sensor Technology Division, 565 Technology Square, 8th floor, Cambridge, MA 02139

SECTION VIII:

FLORIDA ATLANTIC UNIVERSITY

CTC

Communications Technology Center

College of Engineering
Florida Atlantic University
777 Glades Rd.
P.O. Box 3091
Boca Raton, FL 33431-0991
(561) 367-2343; FAX (561) 367-3418

August 13, 1999

Dr. William Hunt
Georgia Institute of Technology
Department of Electrical Engineering
Atlanta, GA 30332-0250

Dear Bill:

A two-sensor camera concept that had been patented and assigned to FAU was proposed to Georgia Tech as the basis of a camera development program. In this program, under subcontract to Georgia Tech, we first simulated the high-definition, progressive camera architecture at 525 lines using a Pulnix camera and a camera purchased on a previous project. These successful results were reported in a paper published in the SMPTE Journal.¹

We then built a breadboard high definition camera mounted on an optical bench. We used two experimental CCD sensors made by Eastman Kodak in the camera. This was the first demonstration in the world of a color camera with 1920 x 1080 pixels progressively scanned at 60 frames per second. This is the progressive SMPTE 274M standard. It has been a long time objective for HDTV for 15 years. This is the first camera to achieve that objective. Since this was a breadboard using an early version of the Eastman CCD, it was not a practical camera. Since it was an optical bench, you couldn't move it.

We subsequently built a more practical camera, on a DARPA contract, using the equipment purchased on this subcontract, plus additional materials. This camera was delivered to the NAVY for underwater experiments.

We built a commercial camera of this general design on a commercial contract and demonstrated it at the annual SMPTE conference in New York City. Because of the considerable interest generated by this demonstration, our commercial sponsor licensed the commercial rights to the design.

We are currently completing an improved version of the camera on DARPA funds for delivery to the Navy. This uses an improved CCD sensor by Eastman Kodak and has all digital processing of the signals. It also can be recorded on a RAID recorder and has a standard SMPTE 292 digital serial output, so that the signal can be transmitted over either coaxial cable or fiber.

Enclosed is the final close out documentation required by the subcontract. Thank you for your assistance in these matters.

Sincerely,



William E. Glenn, Ph.D.
Director

¹ W. E. Glenn, and J. W. Marcinka, "The Development of a 1920 x 1080 Pixel Color CCD Camera Progressively Scanned at 60 Frames Per Second," SMPTE Journal, October 1997.



FLORIDA ATLANTIC UNIVERSITY

777 GLADES ROAD

P.O. BOX 3091

BOCA RATON, FLORIDA 33431-0991

DIVISION OF SPONSORED RESEARCH

(561) 297-2310

FAX (561) 297-2319

July 20, 1998

Ms. Danielle J. Herrmann
Contracting Officer
Georgia Institute of Technology
Atlanta, Georgia 30332-0420

Re: Closeout of Subcontract No. E-21-H83-G4

Dear Ms. Herrmann,

We are submitting the following close out documents per our conversation.

- Subcontractor's Release (State the cumulative amount invoiced)
- Subcontract's Assignment of Refunds, Rebates, Credits, and Other Amounts
- Final Property Report
- Copy of Final Invoice

If you have any administrative questions please contact me at 561-297-2366. Thank you for your assistance in these matters.

Respectfully,

Beth A. Swerdloff, Assistant Director
Division of Sponsored Research

c: J. Engelbrecht
E. Bommel

GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332-0420

Subcontract No. E-21-H83-G4

Final Property Inventory Report

(Check one)

1. X Attached is a listing of property items acquired and/or accountable under this subcontract. Requested disposition of each item is also noted.
2. I certify that there were no items of property acquired and/or accountable under this subcontract.

7/9/98
Date


Signature

Edwin P. Bemmel
Typed Name

Acting Assistant Controller
Title

FLORIDA ATLANTIC UNIVERSITY
INVENTORY REPORT BY DEPARTMENT/PROJECT

SY320BBB SUS
RUN DATE: 07/09/98 TIME: 11:30:34.5

DEPT NAME: ELECTRICAL ENGINEERING
PROJ NAME: DEV OF HIGH DEV SYST CAMERAS (GLENN)
DEPT/PROJ: 1620-057-02
PROJ LEADER:

PROP NO.	P O NUMBER	D E S C R I P T I O N	S E R I A L N U M B E R	I N V D A T E	P Y U N I Q U E	B L D G	R O O M	C O S T	A Q D A T E	C
00119006	E6061100001	SOFTWARE, CAD PROMIXED	408832	100697		0043	SE0330	27000.00	112095	1
00119166	E6067050001	OSCILLOSCOPE: TEKTRONIX	B011563	100697		0043	SE0335	13731.75	122095	1
00119418	E6062580001	SOFTWARE, TANGO PRO	152276	102197		0043	SE0330	6566.25	121495	1
00119634		DIRECT TV RECEIVER & DIS	523344168	92297		0043	SE0330	855.13	10396	1
00119934	E6104150001	SOFTWARE: XILINX	C0039929	100697		0043	SE0330	1275.83	40296	1
00119935	E6104150001	SOFTWARE: XILINX	C0039898	111097		0043	SE0443C	1275.82	40296	1
00121117	E6140990001	CAMERA: VIDEO/PULNIX DEM	000094	100697		0043	SE0330	3646.50	111296	1
00121539	E7071200001	PRINTER: APPLE LASER MOD N/A		92597		0043	SE0486	2116.00	10897	1
00121581	E7071180001	COMPUTER: DELL 6180/OP G 810TB		100697		0043	SE0335	2867.00	12897	1
00122734	E7113990001	OSCILLOSCOPE: 500MHZ 2+2 B020764		72497		0043	SE0335	8935.89	62497	1
00122735	E7113990001	OSCILLOSCOPE: 500MHZ 2+2 TDS520B		100697		0043	SE0335	8935.90	62497	1
00122771	E7113980001	GENERATOR: MOD/#81 PULSE 765645		100697		0043	SE0335	4089.30	70297	1
00122772	E7113980001	GENERATOR: MOD/#81 PULSE 765646		100697		0043	SE0335	4089.30	70297	1
00123098	E7142740001	PC BOARDS P/NM5CN01 2-P N/A		81597	NON-TAGGABLE	0043	SE0356	1105.00	62597	1
00123099	E7142740002	PC BOARDS P/N COMBINER N/A		81597	NON-TAGGABLE	0043	SE0356	1105.00	62597	1
00123199	E7112220001	MONITOR: TV HI-DEFINITION 0604720		102797		0043	SE0335	3000.00	82597	1
00123200	E7112220001	MONITOR: TV HI-DEFINITION 0604721		100697		0043	SE0330	3000.00	82597	1

SUB TOTAL FOR DEPARTMENT: 162005702

-----ACTIVE-----
COUNT COST

17 93,594.67

SY320BBB SUS
RUN DATE: 07/09/98 TIME: 11:30:34.6
DEPT/PROJ: 1620-057-02
PROJ LEADER:

FLORIDA ATLANTIC UNIVERSITY
INVENTORY REPORT BY DEPARTMENT/PROJECT
DEPT NAME: ELECTRICAL ENGINEERING
PROJ NAME: DEV OF HIGH DEV SYST CAMERAS (GLENN)

PAGE NO. 2
EFFECTIVE: 07/09/98

PROP NO.	P O NUMBER	D E S C R I P T I O N	S E R I A L N U M B E R	I N V D A T E	P Y U N I Q U E	B L D G	R O O M	C O S T	A Q D A T E
----------	------------	-----------------------	-------------------------	---------------	-----------------	---------	---------	---------	-------------

* PROPERTY CONDITION CODES
1 - NEW
2 - GOOD
3 - FAIR
4 - POOR
BLANK - NEEDS A CONDITION CODE

***** E N D O F R E P O R T *****

**SUBCONTRACTOR'S ASSIGNMENT OF REFUNDS,
REBATES, CREDITS, AND OTHER AMOUNTS**
SUBCONTRACT No. E-21-H83-G4

Pursuant to the terms of Subcontract No. E-21-H83-G4 under Prime Contract No. NAGW2753 and in consideration of the compensation as provided in the said subcontract and any assignment thereunder, the Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431

(Subcontractor's Name and Address)
hereinafter called the Subcontractor does hereby:

1. Assign, transfer, set over and release to Georgia Institute of Technology (hereinafter called the Contractor), all rights, title and interest to all refunds, rebates, credits and other amounts, (including any interest thereon) arising out of the performance of the said subcontract, together with all the rights of action accrued or which may hereafter accrue thereunder.
2. Agree to take whatever action may be necessary to effect prompt collection of all refunds, rebates, credits, and other amounts (including any interest thereon) due or which may become due, and to promptly forward any such proceeds to Georgia Institute of Technology, Office of Contract Administration, Atlanta, Georgia 30332. The reasonable costs of any such action to effect collection shall constitute allowable costs when approved by the Contracting Officer as stated in the said subcontract and may be applied to reduce any amounts otherwise payable to the Contractor under the terms hereof.
3. Agree to cooperate fully with the Contractor as to any claim or suit in connection with refunds, rebates, credits, and other amounts due (including any interest thereon); to execute any protest, pleading, application, power of attorney, or other papers in connection therewith; and to permit the Contractor to represent it at any hearing, trial, or other proceeding arising out of such claim or suit.

IN WITNESS WHEREOF, this release has been executed this 9th day of July 19 98.

Florida Atlantic University
(Subcontractor)

Signature: _____

Typed Name: Edwin P. Bemmel

Title: Acting Assistant Controller

WITNESSES

Edwin P. Bemmel
Patricia Scarcelli

(Note: In the case of a corporation, witnesses are not required, but the following certificate must be completed.)

CERTIFICATE

I, _____, certify that I am _____ (Official Title)
of the _____ named as Subcontractor in the foregoing release;
that _____ who signed said release on behalf of the Subcontractor,
was then _____ (Official Title) of said _____;
that said release was duly signed for and on behalf of said _____ by
authority of its governing body and is within the scope of its corporate powers.

Signature: _____

Date: _____

Typed Name: _____

Title: _____

SUBCONTRACTOR'S RELEASE
SUBCONTRACT No. E-21-H83-G4

Pursuant to the terms of Subcontract No. E-21-H83-G4 under Prime Contract No. NAGW2753
and in consideration of the sum of Eight Hundred & Seventy-nine Thousand Three Hundred
& Sixty-five and 86 cents dollars (\$ 879,365.86) which has been or is to be paid
under the said subcontract to: E-21-H83-G4
Florida Atlantic University
777 Glades Road, Boca Raton, Florida 33431-0991
(Subcontractor's Name and Address)

(hereinafter called the Subcontractor) or to its assignees, if any, the Subcontractor upon payment of the said sum by Georgia Institute of Technology (hereinafter called the Contractor) does remise, release, and discharge the Contractor, its officers, agents, and employees, of and from all liabilities, obligations, claims and demands whatsoever under or arising from the said subcontract.

The Subcontractor agrees, in connection with patent matters that it will comply with all of the provisions of the said subcontract, including without limitation those provisions relating to notification to the Contracting Officer and relating to the defense of prosecution of litigation.

IN WITNESS WHEREOF, this release has been executed this 9th day of July 19 98.

Florida Atlantic University
(Subcontractor)

Signature: _____

Typed Name: Edwin P. Bemmel

Title: Acting Assistant Controller

WITNESSES

Handwritten Signature 1
Handwritten Signature 2

CERTIFICATE

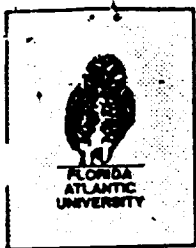
I, _____, certify that I am _____ (Official Title)
of the _____ named as Subcontractor in the foregoing release;
that _____ who signed said release on behalf of the Subcontractor,
was then _____ (Official Title) of said _____;
that said release was duly signed for and on behalf of said _____ by
authority of its governing body and is within the scope of its corporate powers.

Date: _____

Signature: _____

Typed Name: _____

Title: _____



FLORIDA ATLANTIC UNIVERSITY
PO Box 3091
Boca Raton, Florida 33431-0991
(561) 297-2033

INVOICE No. 00700

CONTRACTS & GRANTS INVOICE

DATE: 04/20/

GEORGIA INSTITUTE OF TECHNOLOGY
C/O DR WILLIAM HUNT
DEPARTMENT OF ELECTRICAL ENGINEERING
ATLANTA
GEORGIA 30332-0250

• MAKE CHECKS PAYABLE TO: FLORIDA ATLANTIC UNIVERSITY

• MAIL TO THE CASHIER'S OFFICE AT THE ABOVE ADDRESS

• PLEASE INCLUDE THE INVOICE NUMBER ON YOUR CHECK

AGREEMENT/AWARD	DEBTOR NUMBER	FAU ACCOUNT NO.	OBJECT CODE
NAGW2753	GC0000132	02-1620-057 EX:00	007001

PROJECT TITLE: DEVELOPMENT OF HIGH DEFINITION SYSTEMS CAMERAS
PI: WILLIAM GLENN
FOR THE MONTH ENDING: DECEMBER, 1997

	CURRENT	CUMULATIVE
SALARY:	\$ 28,203.92	\$ 363,737.04
OPS:	0.00	25,341.60
FRINGE:	8,204.16	109,527.64
TRAVEL:	0.00	10,870.99
EXPENSE:	-219.90	87,380.01
OCO:	0.00	57,721.85
TUITION:	413.90	413.90
SUBCONTRACT ABOVE \$25k	0.00	0.00
TOTAL DIRECT COST	\$ 36,602.08	\$ 654,993.03
INDIRECT COST @ 45.00%	16,383.64	224,372.83
TOTAL COST	\$ 52,985.72	\$ 879,365.86
TOTAL DUE THIS INVOICE	\$ 52,985.72	

SINCERELY,

KETIE ST. LOUIS
ACCOUNTING COORDINATOR

PLEASE REMIT PROMPTLY. PAYABLE WITHIN 30 DAYS.
FOR PAYMENTS BY JOURNAL TRANSFER:
SAMAS ACCOUNT NUMBER: 4920 2 655007 48900700 60
CATEGORY CODE: 001500

SECTION VIII:

FLORIDA ATLANTIC UNIVERSITY

CTC

Communications Technology Center

College of Engineering
Florida Atlantic University
777 Glades Rd.
P.O. Box 3091
Boca Raton, FL 33431-0991
(561) 367-2343; FAX (561) 367-3418

August 13, 1999

Dr. William Hunt
Georgia Institute of Technology
Department of Electrical Engineering
Atlanta, GA 30332-0250

Dear Bill:

A two-sensor camera concept that had been patented and assigned to FAU was proposed to Georgia Tech as the basis of a camera development program. In this program, under subcontract to Georgia Tech, we first simulated the high-definition, progressive camera architecture at 525 lines using a Pulnix camera and a camera purchased on a previous project. These successful results were reported in a paper published in the SMPTE Journal.¹

We then built a breadboard high definition camera mounted on an optical bench. We used two experimental CCD sensors made by Eastman Kodak in the camera. This was the first demonstration in the world of a color camera with 1920 x 1080 pixels progressively scanned at 60 frames per second. This is the progressive SMPTE 274M standard. It has been a long time objective for HDTV for 15 years. This is the first camera to achieve that objective. Since this was a breadboard using an early version of the Eastman CCD, it was not a practical camera. Since it was an optical bench, you couldn't move it.

We subsequently built a more practical camera, on a DARPA contract, using the equipment purchased on this subcontract, plus additional materials. This camera was delivered to the NAVY for underwater experiments.

We built a commercial camera of this general design on a commercial contract and demonstrated it at the annual SMPTE conference in New York City. Because of the considerable interest generated by this demonstration, our commercial sponsor licensed the commercial rights to the design.

We are currently completing an improved version of the camera on DARPA funds for delivery to the Navy. This uses an improved CCD sensor by Eastman Kodak and has all digital processing of the signals. It also can be recorded on a RAID recorder and has a standard SMPTE 292 digital serial output, so that the signal can be transmitted over either coaxial cable or fiber.

Enclosed is the final close out documentation required by the subcontract. Thank you for your assistance in these matters.

Sincerely,



William E. Glenn, Ph.D.
Director

¹ W. E. Glenn, and J. W. Marcinka, "The Development of a 1920 x 1080 Pixel Color CCD Camera Progressively Scanned at 60 Frames Per Second," SMPTE Journal, October 1997.



FLORIDA ATLANTIC UNIVERSITY

777 GLADES ROAD

P.O. BOX 3091

BOCA RATON, FLORIDA 33431-0991

DIVISION OF SPONSORED RESEARCH
(561) 297-2310
FAX (561) 297-2319

July 20, 1998

Ms. Danielle J. Herrmann
Contracting Officer
Georgia Institute of Technology
Atlanta, Georgia 30332-0420

Re: Closeout of Subcontract No. E-21-H83-G4

Dear Ms. Herrmann,

We are submitting the following close out documents per our conversation.

- Subcontractor's Release (State the cumulative amount invoiced)
- Subcontract's Assignment of Refunds, Rebates, Credits, and Other Amounts
- Final Property Report
- Copy of Final Invoice

If you have any administrative questions please contact me at 561-297-2366. Thank you for your assistance in these matters.

Respectfully,

A handwritten signature in cursive script, reading "Beth A. Swerdloff".

Beth A. Swerdloff, Assistant Director
Division of Sponsored Research

c: J. Engelbrecht
E. Bommel

GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332-0420

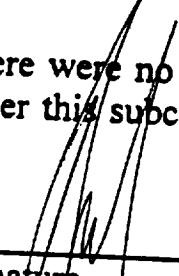
Subcontract No. E-21-H83-G4

Final Property Inventory Report

(Check one)

1. X Attached is a listing of property items acquired and/or accountable under this subcontract. Requested disposition of each item is also noted.
2. I certify that there were no items of property acquired and/or accountable under this subcontract.

7/9/98
Date


Signature

Edwin P. Bemmel
Typed Name

Acting Assistant Controller
Title

FLORIDA ATLANTIC UNIVERSITY
INVENTORY REPORT BY DEPARTMENT/PROJECT

SY3208BB SUS
RUN DATE: 07/09/98 TIME: 11:30:34.5

DEPT NAME: ELECTRICAL ENGINEERING
PROJ NAME: DEV OF HIGH DEV SYST CAMERAS (GLENN)
DEPT/PROJ: 1620-057-02
PROJ LEADER:

PROP NO.	P O NUMBER	D E S C R I P T I O N	S E R I A L N U M B E R	I N V D A T E	P Y U N I Q U E	B L D G	R O O M	C O S T	A Q D A T E	C
00119006	E6061100001	SOFTWARE, CAD PROMIXED	408832	100697		0043	SE0330	27000.00	112095	1
00119166	E6067050001	OSCILLOSCOPE: TEKTRONIX	B011563	100697		0043	SE0335	13731.75	122095	1
00119418	E6062580001	SOFTWARE, TANGO PRO	152276	102197		0043	SE0330	6566.25	121495	1
00119634		DIRECT TV RECEIVER & DIS	523344168	92297		0043	SE0330	855.13	10396	1
00119934	E6104150001	SOFTWARE: XILINX	C0039929	100697		0043	SE0330	1275.83	40296	1
00119935	E6104150001	SOFTWARE: XILINX	C0039898	111097		0043	SE0443C	1275.82	40296	1
00121117	E6140990001	CAMERA: VIDEO/PULNIX DEM	000094	100697		0043	SE0330	3646.50	111296	1
00121539	E7071200001	PRINTER: APPLE LASER MOD N/A		92597		0043	SE0486	2116.00	10897	1
00121581	E7071180001	COMPUTER: DELL 6180/OP G 810TB		100697		0043	SE0335	2867.00	12897	1
00122734	E7113990001	OSCILLOSCOPE: 500MHZ 2+2 B020764		72497		0043	SE0335	8935.89	62497	1
00122735	E7113990001	OSCILLOSCOPE: 500MHZ 2+2 TDS520B		100697		0043	SE0335	8935.90	62497	1
00122771	E7113980001	GENERATOR: MOD/#81 PULSE 765645		100697		0043	SE0335	4089.30	70297	1
00122772	E7113980001	GENERATOR: MOD/#81 PULSE 765646		100697		0043	SE0335	4089.30	70297	1
00123098	E7142740001	PC BOARDS P/NM5CN0N1 2-P N/A		81597	NON-TAGGABLE	0043	SE0356	1105.00	62597	1
00123099	E7142740002	PC BOARDS P/N COMBINER N/A		81597	NON-TAGGABLE	0043	SE0356	1105.00	62597	1
00123199	E7112220001	MONITOR: TV HI-DEFINITION 0604720		102797		0043	SE0335	3000.00	82597	1
00123200	E7112220001	MONITOR: TV HI-DEFINITION 0604721		100697		0043	SE0330	3000.00	82597	1

SUB TOTAL FOR DEPARTMENT: 162005702

-----ACTIVE-----
COUNT COST

17 93,594.67

**SUBCONTRACTOR'S ASSIGNMENT OF REFUNDS,
REBATES, CREDITS, AND OTHER AMOUNTS**
SUBCONTRACT No. E-21-H83-G4

Pursuant to the terms of Subcontract No. E-21-H83-G4 under Prime Contract No. NAGW2753 and in consideration of the compensation as provided in the said subcontract and any assignment thereunder, the Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431

(Subcontractor's Name and Address)

hereinafter called the Subcontractor does hereby:

1. Assign, transfer, set over and release to Georgia Institute of Technology (hereinafter called the Contractor), all rights, title and interest to all refunds, rebates, credits and other amounts, (including any interest thereon) arising out of the performance of the said subcontract, together with all the rights of action accrued or which may hereafter accrue thereunder.
2. Agree to take whatever action may be necessary to effect prompt collection of all refunds, rebates, credits, and other amounts (including any interest thereon) due or which may become due, and to promptly forward any such proceeds to Georgia Institute of Technology, Office of Contract Administration, Atlanta, Georgia 30332. The reasonable costs of any such action to effect collection shall constitute allowable costs when approved by the Contracting Officer as stated in the said subcontract and may be applied to reduce any amounts otherwise payable to the Contractor under the terms hereof.
3. Agree to cooperate fully with the Contractor as to any claim or suit in connection with refunds, rebates, credits, and other amounts due (including any interest thereon); to execute any protest, pleading, application, power of attorney, or other papers in connection therewith; and to permit the Contractor to represent it at any hearing, trial, or other proceeding arising out of such claim or suit.

IN WITNESS WHEREOF, this release has been executed this 9th day of July 19 98.

Florida Atlantic University

(Subcontractor)

Signature: _____

Typed Name: Edwin P. Bemmel

Title: Acting Assistant Controller

WITNESSES

Pauline W. Clark
Patricia Scarcelli

(Note: In the case of a corporation, witnesses are not required, but the following certificate must be completed.)

CERTIFICATE

I, _____, certify that I am _____ (Official Title)
of the _____ named as Subcontractor in the foregoing release;
that _____ who signed said release on behalf of the Subcontractor,
was then _____ (Official Title) of said _____;
that said release was duly signed for and on behalf of said _____ by
authority of its governing body and is within the scope of its corporate powers.

Signature: _____

Typed Name: _____

Title: _____

Date: _____

SUBCONTRACTOR'S RELEASE
SUBCONTRACT No. E-21-H83-G4

Pursuant to the terms of Subcontract No. E-21-H83-G4 under Prime Contract No. NAGW2753
and in consideration of the sum of Eight Hundred & Seventy-nine Thousand Three Hundred
& Sixty-five and 86 cents dollars (\$ 879,365.86) which has been or is to be paid
under the said subcontract to: E-21-H83-G4
Florida Atlantic University
777 Glades Road, Boca Raton, Florida 33431-0991
(Subcontractor's Name and Address)

(hereinafter called the Subcontractor) or to its assignees, if any, the Subcontractor upon payment of the said sum by Georgia Institute of Technology (hereinafter called the Contractor) does remise, release, and discharge the Contractor, its officers, agents, and employees, of and from all liabilities, obligations, claims and demands whatsoever under or arising from the said subcontract.

The Subcontractor agrees, in connection with patent matters that it will comply with all of the provisions of the said subcontract, including without limitation those provisions relating to notification to the Contracting Officer and relating to the defense of prosecution of litigation.

IN WITNESS WHEREOF, this release has been executed this 9th day of July 19 98.

Florida Atlantic University
(Subcontractor)

Signature: _____

Typed Name: Edwin P. Bemmel

Title: Acting Assistant Controller

WITNESSES

Handwritten Signature
Patricia Scavelli

CERTIFICATE

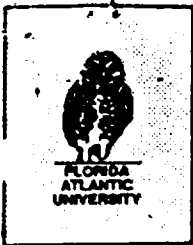
I, _____, certify that I am _____ (Official Title)
of the _____ named as Subcontractor in the foregoing release;
that _____ who signed said release on behalf of the Subcontractor,
was then _____ (Official Title) of said _____;
that said release was duly signed for and on behalf of said _____ by
authority of its governing body and is within the scope of its corporate powers.

Signature: _____

Date: _____

Typed Name: _____

Title: _____



FLORIDA ATLANTIC UNIVERSITY
PO Box 3091
Boca Raton, Florida 33431-0991
(561) 297-2033

INVOICE No. 00700

CONTRACTS & GRANTS INVOICE

DATE: 04/20/

GEORGIA INSTITUTE OF TECHNOLOGY
C/O DR WILLIAM HUNT
DEPARTMENT OF ELECTRICAL ENGINEERING
ATLANTA
GEORGIA 30332-0250

* MAKE CHECKS PAYABLE TO: FLORIDA ATLANTIC UNIVERSITY
* MAIL TO THE CASHIER'S OFFICE AT THE ABOVE ADDRESS
* PLEASE INCLUDE THE INVOICE NUMBER ON YOUR CHECK

AGREEMENT/AWARD	DEBTOR NUMBER	FAU ACCOUNT NO.	OBJECT CODE
NAGW2753	GC0000132	02-1620-057 EX:00	007001

PROJECT TITLE: DEVELOPMENT OF HIGH DEFINITION SYSTEMS CAMERAS
PI: WILLIAM GLENN
FOR THE MONTH ENDING: DECEMBER, 1997

	CURRENT	CUMULATIVE
SALARY:	\$ 28,203.92	\$ 363,737.04
OPS:	0.00	25,341.60
FRINGE:	8,204.16	109,527.64
TRAVEL:	0.00	10,870.99
EXPENSE:	-219.90	87,380.01
OCO:	0.00	57,721.85
TUITION:	413.90	413.90
SUBCONTRACT ABOVE \$25k	0.00	0.00
TOTAL DIRECT COST	\$ 36,602.08	\$ 654,993.03
INDIRECT COST @ 45.00%	16,383.64	224,372.83
TOTAL COST	\$ 52,985.72	\$ 879,365.86
TOTAL DUE THIS INVOICE	\$ 52,985.72	

SINCERELY,

KETIE ST. LOUIS
ACCOUNTING COORDINATOR

PLEASE REMIT PROMPTLY. PAYABLE WITHIN 30 DAYS.
FOR PAYMENTS BY JOURNAL TRANSFER:
SAMAS ACCOUNT NUMBER: 4920 2 655007 48900700 60
CATEGORY CODE: 001500